# DOCUMENT RESUME

| ED 047 283 | 24 | AC 010 116 |
|---|---|---|

| | |
|---|---|
| AUTHOR | Grotelueschen, Arden; McQuarrie, Duncan |
| TITLE | 1970 Adult Education Research Conference (AERC) Evaluation Workshop: An Evaluative Report. |
| SPONS AGENCY | Office of Education (DHEW), Washington, D.C. Bureau of Research. |
| BUREAU NO | BR-0-8006 |
| PUB DATE | Jun 70 |
| GRANT | OEG-0-70-2844(520) |
| NOTE | 171p. |
| EDRS PRICE | EDRS Price MF-$0.65 HC-$6.58 |
| DESCRIPTORS | *Adult Education, Changing Attitudes, *Conferences, Educational Objectives, Evaluation Techniques, Measurement Instruments, Orientation, *Program Evaluation, *Research, Tests, *Workshops |
| IDENTIFIERS | *Adult Education Research Conference |

ABSTRACT

Major objectives of this Adult Education Research Conference workshop were sixfold: to examine in detail the Stake Countenance Model of Evaluation as a potential conceptual framework for adult education evaluation; practice using it to identify and categorize relevant variables and relationships; design evaluation plans for typical adult education programs; understand different types of evaluation procedures more fully; compare two basic styles of inquiry (research and evaluative); and ascertain the function of communication in evaluation. It was planned that participants should complete an attitude scale covering orientations (research, service, teaching, objectives, judgment, degree of confidence) toward evaluation, then pursue a simulation exercise and other problem solving activities. Intended instructional activities were largely carried out. The reported usefulness of a specially prepared instructional resource (the Evaluation Notebook) more than compensated, both during and after the workshop, for the lack of the preliminary activities originally planned; and the attitude scale was well received. Significant increases in knowledge, understanding, and application of the model were also noted, together with much participant satisfaction with outcomes. (LY)

BR 0-8006
PA 24
AC

ED0 47283

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

# FINAL REPORT

Project No. 0-8006

Grant No. OEG-0-70-2844(520)

# 1970 ADULT EDUCATION RESEARCH CONFERENCE (AERC) EVALUATION WORKSHOP: AN EVALUATIVE REPORT

**Arden Grotelueschen and Duncan McQuarrie**
**Center for Instructional Research and Curriculum Evaluation**
**270 Education Building**
**University of Illinois, Urbana, Illinois 61801**

June 1970

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

AC010116

# FINAL REPORT

Project No. 0-8006

Grant No. OEG-0-70-2844(520)

# 1970 ADULT EDUCATION RESEARCH CONFERENCE (AERC) EVALUATION WORKSHOP: AN EVALUATIVE REPORT

Arden Grotelueschen and Duncan McQuarrie
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

June 1970

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

TO

Grandmentor Bob

# TABLE OF CONTENTS

Page

I. Objectives of the Evaluation . . . . . . . . . . . . . . . . . 1

A. Audiences . . . . . . . . . . . . . . . . . . . . . . . 1

B. Anticipated Decisions . . . . . . . . . . . . . . . . . 1

II. Specifications of the Workshop . . . . . . . . . . . . . . . 2

A. Educational Philosophy and Subject Matter of the Workshop . . . . . . . . . . . . . . . . . . . . 2

B. Learning Objectives, Staff Aims . . . . . . . . . . . . . 2

C. Instructional Procedures, Tactics, Media . . . . . . . . . 3

D. Participants and Staff . . . . . . . . . . . . . . . . . 6

E. Instructional Setting . . . . . . . . . . . . . . . . . 7

F. Standards, Bases for Judging Quality . . . . . . . . . . . 7

III. Program Outcomes . . . . . . . . . . . . . . . . . . . . . 9

A. Opportunities and Experiences Provided . . . . . . . . . . 9

B. Participant Outcomes . . . . . . . . . . . . . . . . . . 9

1. Tangible . . . . . . . . . . . . . . . . . . . . . 9

2. Cognitive . . . . . . . . . . . . . . . . . . . . . 10

3. Affective . . . . . . . . . . . . . . . . . . . . . 12

4. Social and Collegial . . . . . . . . . . . . . . . . 15

C. Side Effects and Bonuses . . . . . . . . . . . . . . . . 16

D. Costs . . . . . . . . . . . . . . . . . . . . . . . . . 16

IV. Congruencies . . . . . . . . . . . . . . . . . . . . . . . 17

A. Antecedents . . . . . . . . . . . . . . . . . . . . . . 17

1. The Setting and Facilities . . . . . . . . . . . . . 17

2. Objectives . . . . . . . . . . . . . . . . . . . . . 17

3. Participants . . . . . . . . . . . . . . . . . . . . 19

TABLE OF CONTENTS (Continued)

Page

B. Transactions . . . . . . . . . . . . . . . . . . . . . . . 20

C. Outcomes . . . . . . . . . . . . . . . . . . . . . . . . . 21

V. Judgments of Worth . . . . . . . . . . . . . . . . . . . . . 23

A. Value of Outcomes . . . . . . . . . . . . . . . . . . . . . 23

B. Relevance of Objectives to Needs . . . . . . . . . . . . . 23

C. Relation to Standards . . . . . . . . . . . . . . . . . . . 24

D. Usefulness of Evaluation Information Gathered . . . . . . . 24

VI. Appendices

A. Tentative Outline of Workshop Activities

B. CIRCE Attitude Scale

C. Floor Plan of Evaluation Workshop Meeting Area

D. AERC Evaluation Workshop Notebook

E. AERC Evaluation Workshop Case Studies

F. Countenance Model Quiz

G. Participant Summative Evaluation Form

H. AERC Evaluation Workshop Participant Information

I. 1970 Adult Education Research Conference Participant Roster

J. AERC Evaluation Workshop Schedule

K. AERC Evaluation Workshop Instructional Staff

L. 1970 AERC Evaluation Workshop Observation Guide

# I. OBJECTIVES OF THE EVALUATION

## A. Audiences

This report is intended for the use of five primary audiences: the funding source *(Research Training Branch, Bureau of Research, U. S. Office of Education, Department of Health, Education and Welfare)*; the sponsoring organization *(Adult Education Research Conference Executive Committee of 1970 and 1971)*; the unit providing the training staff *(Center for Instructional Research and Curriculum Evaluation, College of Education, University of Illinois at Urbana-Champaign)*; the participants in the workshop *(1970 Adult Education Research Conference Participants)*; and the agency responsible for channeling the external funding to the sponsoring organization *(Division of University Extension, Special Programs and Research, University of Illinois)*.

## B. Anticipated Decisions

Of the above audiences, it is anticipated that three audiences might most likely use this report in making future decisions. The *Research Training Branch* might use this report in making determinations about the usefulness of continuing to support evaluation training for members of specialized professional fields. The staff of *CIRCE* might use this report as one piece of evidence in their decision to continue to periodically offer short-term evaluation workshops. The *AERC* Executive Committee might use the results of this evaluation in designing future annual conferences.

## II. SPECIFICATIONS OF THE WORKSHOP

### A. Educational Philosophy and Subject Matter of the Workshop

In earlier years the *Adult Education Research Conference*, formerly known as the *National Seminar on Adult Education Research*, regarded in-service training in research and evaluation as a major purpose for conducting its annual meetings. In more recent years this opportunity for continued professional development has not been capitalized upon as in earlier years. More time was being devoted to the findings of research and evaluative studies, and little time was devoted to conceptual and methodological aspects of research and evaluation in adult education. In an attempt to return to the earlier emphasis, the Executive Committee of the 1970 Conference decided to devote a major portion of the annual Conference to formalized training by conducting a workshop on educational evaluation.

In the past the participants at the annual Conference have had major research and evaluation responsibilities related to adult education. Although this group has learned by experience the merits of evaluation, it was believed by the Executive Committee that most adult educators had not kept up with the recent writings of evaluation theorists such as Cronbach, Popham, Scriven, Stake, and Stufflebeam. It was decided by the workshop staff to emphasize the development of a *single* relevant theoretical framework from which adult education specialists might approach problems of program evaluation. While they acknowledged the existence of alternative theoretical positions the workshop staff felt that given the limited amount of instructional time only one evaluation model could be taught well enough to serve as a frame of reference. Since eight of the nine proposed staff members for the workshop were affiliated with the *Center for Instructional Research and Curriculum Evaluation* (CIRCE) the evaluation model chosen (the Stake Countenance Model) was the one associated with that Center.

### B. Learning Objectives, Staff Aims

The intended general purpose was to present a short-term, highly intensive workshop designed to provide experiences that would broaden the conceptual base of adult education researchers and evaluators so that they would have a more relevant theoretical framework from which to approach problems of program evaluation. The intended major instructional objectives of the workshop, in order of their priority, were:

1. To examine in detail the Stake Countenance Model of Evaluation in order to provide a conceptual framework for adult education evaluation.

2. To practice using the Stake Countenance Model for the identification and categorization of relevant evaluation variables and relationships.

3. To design evaluation plans for typical adult education programs.

4. To better understand the distinction between summative and formative evaluation procedures.

5. To compare and contrast research and evaluative styles of inquiry.

## C. Instructional Procedures, Tactics, Media

The design of the instructional program was guided by theoretical considerations, logical analysis, and past experience in conducting evaluation workshops. What follows is a description of the activities (transactions) proposed in the statement submitted November 1, 1969 to the U. S. Office of Education. In a subsequent section the congruency between the proposed and the actual transactions will be described.

It was planned that 150 participants would be divided into three Instructional Groups. This arrangement, it was believed, would offer the opportunity to give differential treatments to randomized groups or test the effects of differential grouping. Certain information about potential workshop participants was to be required prior to their possible acceptance into the workshop. Data on such aspects as field of adult education, background in evaluation, biographical characteristics, and attitude toward evaluation were to be available to the staff prior to the convening of the workshop and it was planned that selected participants could be blocked on selected variables and randomly assigned to groups.

For each group there was to be a corresponding Instructional Team composed of two instructors (one focused on evaluation and one on adult education) and one graduate assistant. It was believed that the probability of relating evaluation to adult education would be increased by the fact that the staff members related well with each other personally and professionally, and that several of the proposed instructors had considerable expertise in both evaluation and adult education. While the groups were to remain in fixed instructional settings or locations, it was planned that the three teams would move from group to group, thus giving the participants a coordinated exposure to all nine evaluation/adult education specialists.

Prior to the workshop, certain instructional activities were to take place. In early February, all of the selected participants were to receive a copy of Stake's Countenance Model, "The Countenance of Educational

Evaluation".[1] All participants were to have read this article prior to their arrival at the workshop. Various subtle and humorous (but sincere) means of encouraging participants to accomplish this task were planned. During the first day of the Conference (two days prior to the workshop)[2] each workshop participant in attendance was to receive a copy of a summary of the Stake Countenance Model. (Persons not attending the Conference, but who had been selected for participation in the workshop, were to be mailed these materials a week before the workshop.) This summary was to be a simple, clear, inclusive outline of the more elaborate model. It was hoped that this redundancy, together with an "Evaluation Vocabulary List" passed out at the same time, would begin to form a stable, discriminable "cognitive structure" that would make the training sessions more productive and learning easier. Additional organizational devices planned during this period were to include a short paper on the history of evaluation and a series of short (3 to 4 minute) professionally-developed tapes on several select topics, e.g., "Outline of Countenance Model," "Distinction Between Research and Evaluation," "Formative and Summative Evaluation," and "The Educational Objectives Controversy." The tapes were to be of the cassette-type with players located in a variety of locations designed to create a novelty effect so as to increase the likelihood of their being played.

The workshop was planned to begin on the third day of the Conference and the tentative outline of the proposed activities is provided in Appendix A. The morning of the first day was designed as an introduction to the field of program evaluation. The first of a series of workshop formative evaluation activities was planned to take place just prior to lunch. These were designed to provide the teams with feedback so that content could be modified (if necessary) in order to better reach the workshop goals. The orientations toward evaluation were to be measured by the CIRCE Attitude Scale (see Appendix B). This instrument was to be completed by the participants prior to the workshop and individual profiles were to be developed and distributed for use during the first session. This instrument had been found useful during prior workshops in eliciting discussion about the implications of differences in attitude toward evaluation. The larger part of the afternoon was to be devoted to a detailed study of the Stake Countenance Model. It was planned that this topic would be introduced by a short presentation on the general role of evaluation models.

The last part of the afternoon session was to be used for the introduction of specific methodological considerations together with a discussion of their relationship to the Stake Countenance Model. Each group was to hear the same topics but in counter-balanced orders so that the strengths of particular members of the instructional staff could be used more than once, i.e., a specialist in a particular methodology could give his presentation to all groups. These sessions were continued as the first part of the next morning's program.

---

[1] Reprint from Teachers College Record, 68, 1967, 523-540.

[2] February 27 and 28 were reserved for the usual Conference activities of paper reading and symposium sessions with March 1 and 2 exclusively set aside for the evaluation workshop activities.

An evening session was scheduled for the first day of the workshop during which time the staff planned on going through a guided simulation exercise in the use of the Stake Countenance Model as a conceptual framework. For this exercise it was planned that video-tapes would be prepared depicting a meeting between a consultant on program evaluation and an adult education program director. It was hoped that three different tapes for different adult education content areas could be developed. It was anticipated that such a dialogue would provide a wide range of data that could be classified by the various categories and relationships of the Stake Countenance Model. While viewing the tape, it was planned that the participants would be asked to identify and record on a provided handout a variety of evaluation variables and relationships. It was further planned that on three different occasions the tapes would be stopped and examples of how evaluation specialists had completed the same task would be distributed. These examples would act as feedback as well as hints and it was thought that they would stimulate discussion of various practical applications of the model.

A second problem-solving session was planned during the second morning. Five adult education case studies corresponding to adult education areas of evaluation were to be prepared. For each of these case studies, it was planned that participants within each Instructional Group would be assigned to one of five small groups representing each case study. The participants were to be assigned so as to maximize the diversity within each group with respect to evaluation frame of reference as measured by the CIRCE Attitude Scale.

It was anticipated that after working on the evaluation designs for approximately one hour, the three small groups working on a common case study across Instructional Groups (one small group from each of the three Instructional Groups) would be combined as a single group. These five composite groups, with approximately 30 participants each, would then compare and discuss similarities and differences in their design to a common case study. It was anticipated that the designs of these case studies would serve as a major means for evaluation of the workshop (Evaluation No. 4). Following lunch the workshop staff planned to present a panel discussion of the case study evaluation designs and answer questions related to them.

The next block of time was left open to allow the staff to cover material not anticipated but indicated as desirable by formative evaluation. The final session was planned to be used to cover the usual administrative details as well as to collect the final set of workshop evaluation data.

## D. Participants and Staff

It was anticipated that the participants selected for the proposed workshop would be professional adult educators having major research and evaluation responsibilities. Because the proposed workshop was to be conducted in conjunction with the annual Adult Education Research Conference, it was thought that there would be considerable overlap between the anticipated 200 Conference attendees and the 150 workshop participants. It was assumed that some Conference attendees would not be able to participate in the workshop, and some of the person not attending the Conference would participate in the workshop. This was to allow for a more representative group of workshop participants with regard to purposes, agency affiliation, and section of the country.

In light of the characteristics of past Conference participants, it was anticipated that participants would reflect the various areas of adult education in which evaluation is conducted. It was thought that these areas would include adult basic education, residential adult education, extension education, and professional education. It was planned that attempts would also be made to recruit persons from the different agencies which sponsor adult education activities (e.g., religious organizations, state departments of education, federal agencies, public schools, universities, and private industry).

Past data suggested that the staff could expect about half the participants to have a doctorate and most of the rest to have a masters degree; about half the participants to be spending at least 25 percent of their time on research and evaluation; and 75 percent of the participants to have completed two or more research projects. It was believed that these adult educators would be among those most likely to influence evaluation practices in the field of adult education.

The proposed staff for the workshop included specialists in both evaluation and adult education. Of the nine staff members proposed for the workshop, eight were to be affiliated with CIRCE at the Urbana-Champaign campus of the University of Illinois. The proposed staff is listed below:

| | |
|---|---|
| Director: | Arden Grotelueschen, Assistant Professor, CIRCE, University of Illinois, Urbana-Champaign |
| Instructors: | Terry Denny, Associate Professor and Research Director of EPIE, CIRCE, University of Illinois, Urbana-Champaign |
| | Douglas Sjogren, Professor, CIRCE, University of Illinois, Urbana-Champaign |
| | Robert Stake, Professor and Associate Director of CIRCE, University of Illinois, Urbana-Champaign |

Resource Persons in Adult Education and/or Evaluation:

Arden Grotelueschen, Assistant Professor,
CIRCE, University of Illinois, Urbana-Champaign

Alan B. Knox, Professor of Adult Education,
Teachers College, Columbia University, New York

Duncan McQuarrie, Research Assistant, CIRCE,
University of Illinois, Urbana-Champaign

Graduate Assistants (All from CIRCE, University of Illinois, Urbana-Champaign):

Dennis Gooler

Margaret Pjojian

Gary Storm

## E. Instructional Setting

It was planned that the workshop be held at the Holiday Inn Central, Nicollet and 13th Street, Minneapolis, Minnesota. The staff anticipated that all workshop attendees would stay at the Holiday Inn Central and preliminary arrangements were to be made on that basis. The workshop activities were to be scheduled for the top floor (14th floor) of the hotel in the dining room and a series of smaller banquet rooms. Appendix C shows a floor plan of the fourteenth floor. The basic plan was to use the "Starlite Room" as one instructional center and to make Rooms 8 and 9 into another, while Rooms 6 and 7 would be a third. Each meeting room was to be set-up school style and provided with blackboard, speakers stand and audio-visual devices upon request.

The area outside of these three rooms was to be used as a serving area for catered coffee breaks scheduled in the mornings and afternoons.

## F. Standards, Bases for Judging Quality

Judgments about the worth of educational programs can be based upon a variety of standards -- desired quantities or qualities of educational programs cited by some authority figure or document. Most frequently educational standards are implicit in the judgments made about an educational program, rarely are they explicitly stated. It was anticipated that the worth of the proposed workshop would be judged with respect to the following explicit standards.

1. Guidelines governing proposals submitted to the Research Training Branch, Bureau of Research, U. S. Office of Education.

2. The statements about what the program of the annual AERC ought to be by leaders in the field of adult education.

3. The number of attendees at the 1969 annual Conference.

4. The regional representation of the participants at the 1969 annual Conference.

# III. PROGRAM OUTCOMES

## A. Opportunities and Experiences Provided

The opportunity to participate and the experience of participating in an educational program might be viewed as important outcomes in and of themselves. These outcomes should be described in an evaluation report. In the evaluation workshop the following general opportunities and experiences occurred: CIRCE staff members broadening their professional perspectives by interacting with adult education researchers, adult educators interacting with evaluation specialists, and both adult educators and evaluation specialists relating and interacting with members of their own groups. To illustrate, adult education researchers were confronted with conceptual issues in evaluation and evaluation specialists were confronted with relating these issues to the practical concerns of the adult educators. The resultant discussions provided an experience for the adult educators unlike those obtained by associations with colleagues in adult education and by reading the professional literature on evaluation. Members of the evaluation staff were provided the opportunity to be exposed to practitioners' questions about evaluation and to observe other staff members attempt to relate to the practical queries raised by adult educators. Finally, the opportunity for staff and participants to become acquainted personally with leaders in the respective fields represented by these groups is an unequaled opportunity and experience.

## B. Participant Outcomes

There are many intended gains in a workshop experience. Most frequently these gains are associated with participant or learner outcomes. There are losses too. They are frequently unintended, except where obvious trade-offs for intended gains have been anticipated.

Following are listed those participants gains and losses that were observed in our evaluation efforts. The categories are arbitrary and not mutually exclusive.

### 1. Tangible

Most of the tangible gains are the instructional materials which participants received during the workshop. Of these, the most significant gain was the AERC Evaluation Workshop Notebook (see Appendix D). This Notebook was viewed by three-fourths of the participants as "extremely" satisfactory. It was also one of the major factors cited by participants as contributing to the successful accomplishment of workshop objectives. In addition, many participants indicated that upon returning to their institutions they used the contents of the Notebook (especially the references and readings) in courses, in-service training seminars, and self-study.

The CIRCE Attitude Scale (Appendix B) was a second tangible gain obtained by the participants. It too was regarded as a satisfactory instructional device, but to a lesser extent than the Notebook. Six participants requested multiple copies or sought permission to reproduce the Scale for use in their instructional activities.

Finally, the workshop library (see references in Notebook for items contained in the library) was a tangible gain for 13 participants, since they won library items raffled off at the end of the workshop. Because a few persons felt everyone in the workshop should have received a library item (a somewhat unreasonable expectation in view of the costs), their not receiving a library item was reported as a loss.

Other tangible gains and losses no doubt occurred. For example, the case studies (see Appendix E) developed for the workshop have multiple uses.

## 2. Cognitive

There are many cognitive outcomes associated with participant learning--for example, increased knowledge, understanding, and application. We has assessed some of the many cognitive outcomes and have gathered participant judgments about them.

Because the workshop was designed around the Stake Countenance Model, the evaluators felt that the participants should be tested on their comprehension of the Model after they had received instruction about it. The participants' performance on a 10 item quiz designed to measure aspects of the Model (see Appendix F for a copy of the quiz) clearly indicated that the participants had knowledge and understanding of the Model. The average score (mean) for the group was 7.2 with the scores ranging from 4 to 10. The score of 3.5 would have been expected of participants without prior knowledge of the topic. It is unlikely that the observed score was due entirely to prior knowledge about the subject since most of the participants indicated before the workshop that they were unfamiliar with the writings of Stake.

Further evidence of cognitive gains was manifest during the discussion of the Model and the application of the Model to Case No. 6 which followed the formal presentation of the Model. The questions asked and the explanations and comments offered by participants were judged by members of the staff to exhibit a substantial level of knowledge and understanding about the Model.

More important than the observed increase in knowledge and understanding that occurred as a result of the workshop experience was the discussion, extension, and application of ideas gained there. Seventy-six percent (32) of the participants responding to a questionnaire (see Appendix G for a copy of this instrument) administered one month after the workshop indicated they had built upon what they had learned in the workshop. Twenty-one percent reported maintaining the ideas presented in the workshop. Only three percent felt they had reverted to their pre-workshop ideas about evaluation.

When participants were asked the extent to which they believed the workshop contributed to their increased competence in approaching and conducting evaluation studies, all respondents felt at least "somewhat" more competent. Thirty-two percent felt "quite" competent and five percent felt "highly" competent.

Eighty-five percent of the participants responding to the questionnaire indicated that they used at least "moderately" what they had learned in the workshop. A majority of the participants who used what they had learned did so in the design and conduct of an evaluation activity. Others indicated applying what they had received in a teaching activity. Still others used what they had learned to assist others in a consultative role.

Almost all the respondents (95%) indicated that they had read or discussed with someone various aspects of evaluation since participating in the workshop. The nature of this involvement was extensive. Several participants reported reviewing the notes and materials of the workshop, but most participants reported discussing the content of the workshop with supervisors, colleagues, and graduate students. Presentations to in-service seminars and staff meetings were frequently mentioned. One person responded that the "materials were placed in a lending library and they were used."

The above information on the cognitive outcomes of the workshop provides evidence to support the conclusion that the workshop was highly successful. Because cognitive outcomes of participants are frequently regarded as the sole criterion for a successful workshop, this conclusion is of even greater significance.

3. Affective

Many positive feelings about the workshop could be inferred from a variety of data. The positive ratings given by participants to various aspects of the workshop (e.g., instruction, materials, content), their observed interactions and interest during the workshop, and their reported continued activity after the workshop are but a few prominent examples.

Direct evidence of attitude change is provided by results obtained from pre- and post-workshop administrations of the CIRCE Attitude Scale. (It will be recalled that this scale is a 49-item inventory designed to reflect an individual's orientation to evaluation.) The results of the pre- and post-workshop administrations indicate that participant attitudes toward evaluation changed during the workshop. More significantly, they changed consistently toward those held by the instructional staff. That is, at the end of the workshop participants valued less highly *research* and *objectives* orientations to evaluation and valued more highly *service, teaching,* and *judgment* orientations. See Figure 1 for a comparison of pre- and post-workshop mean score profiles of participants on attitudes toward evaluation. Table 1 provides summary information of participant performance portrayed in Figure 1.

TABLE 1

Means and Standard Deviations of
Evaluation Orientations by Scale Administration

| | Administration | | | |
|---|---|---|---|---|
| Orientation | Pre-Workshop (*N* = 62) | | Post-Workshop (*N* = 44) | |
| | $\overline{X}$ | S.D. | $\overline{X}$ | S.D. |
| Research | 4.62 | 1.76 | 3.32 | 1.46 |
| Service | 5.79 | 1.39 | 7.05 | 1.88 |
| Teaching | 5.62 | 2.31 | 7.36 | 2.05 |
| Objectives | 4.66 | 2.32 | 3.39 | 2.16 |
| Judgment | 7.54 | 2.04 | 8.30 | 2.00 |
| Confidence | 9.06 | 1.26 | 9.20 | 1.62 |

**A RESEARCH orientation to Evaluation**

The person high on this scale appears to believe that evaluation should rely on precise measurement and statistical analysis to gain *general understanding* of why programs do or do not succeed.

**A SERVICE orientation to Evaluation**

The person high on this scale appears to believe that evaluation should be designed according to the *needs of educators* involved so as to aid them in their present work and future decisions.

**A TEACHING orientation to Evaluation**

The person high on this scale appears to believe that evaluation should be focused considerably on the *quality of teaching* and should discover the intrinsic merit in facilities and in instruction.

**OBJECTIVES orientation to Evaluation**

The person high on this scale appears to believe that instruction, and therefore evaluation, should be focused considerably on apriori statements of objectives, that the merit of the program is largely indicated by the *success of students* in reaching these objectives.

**A JUDGMENT orientation to Evaluation**

The person high on this scale appears to believe that educational evaluation is largely a matter of establishing the worth of the program for various purposes *as perceived* by various groups of persons in and around the program.

**CONFIDENCE in Evaluation**

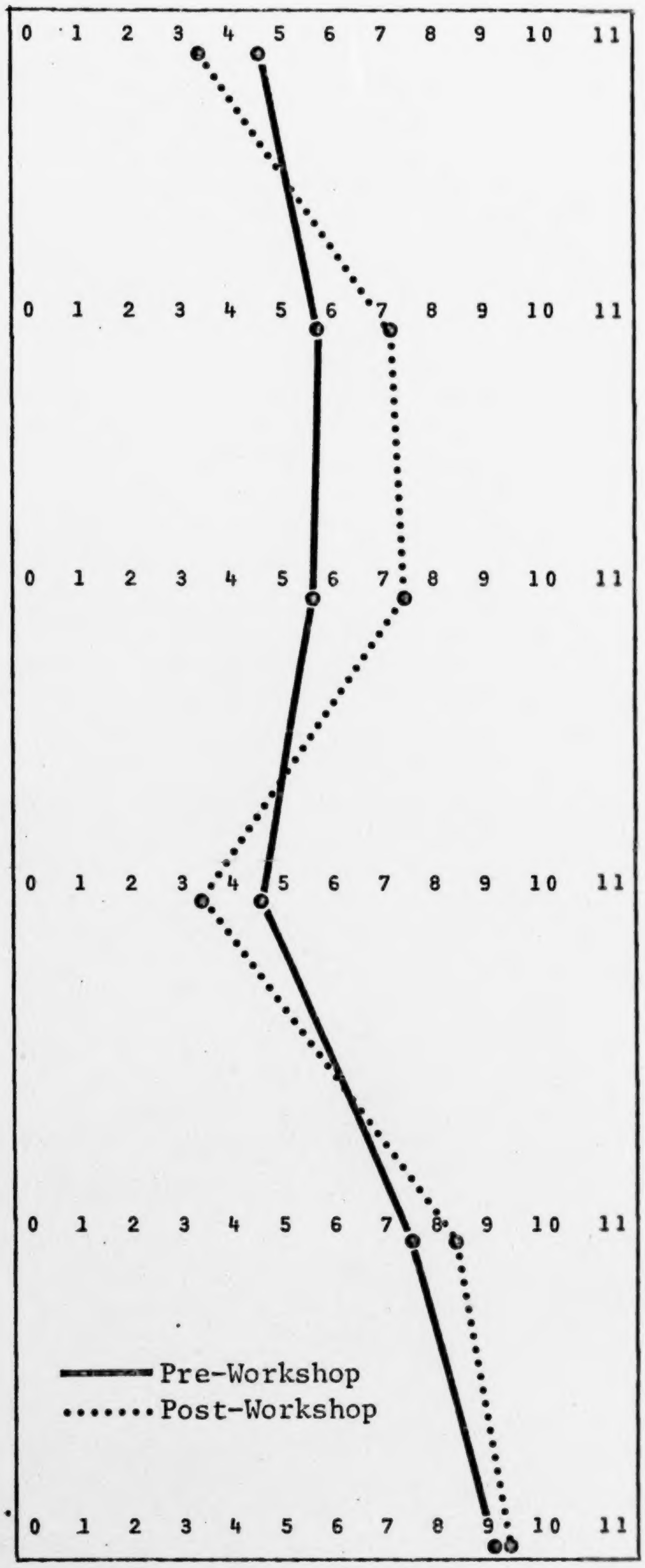


Figure 1. Comparison of pre- and post-workshop mean score profiles of participants on CIRCE Attitude Scale.

The apparent influence of instruction on participant attitude change is considered by the staff to be a very significant outcome of the workshop. Such direct evidence of instructional impact is rarely documented in educational activities.

Further evidence of the positive affective outcomes of the workshop were found in unsolicited testimonials given by participants. Following are several examples of feedback received by the workshop director:

> "The workshop on evaluation contributed much to the positive emotional climate of the Conference. . . . The workshop organized and run by . . . provided a stimulating and informative program."
>
> Reported in Adult Leadership

> "The 1970 AERC was one of the better conferences which I have been priviledged to attend."
>
> Letter to Director

> "I want to take this opportunity to express my sincere appreciation to you and the other CIRCE staff members for an outstanding research conference and evaluation workshop. It was the most informative one that I have attended."
>
> Letter to Director

> "I was very pleased with the experience. I think the threat-free helping, low pressure environment you and your CIRCE colleagues established at the outset was tremendous. I learned. My attitude toward evaluation changed. I was happy!"
>
> Letter to Director

> "A great experience!"
>
> Personal Communication to Director

### 4. Social and Collegial

The social and collegial outcomes of participating in an educational conference are frequently valued highly by participants of that activity. Old acquaintances are renewed--former advisees talk with their mentors, former colleagues talk over old times and new opportunities. New acquaintances are also made--graduate students meet distinguished faculty in the field, practitioners meet theorists. But more importantly; ideas are exchanged, convictions are challenged, allegiances are formed, opportunities are opened, commitments are made, and cooperation is elicited.

To be sure, the preceding occurred at the evaluation workshop. Not all were experienced by everyone; but everyone experienced some social and/or collegial benefits. To document this assertion a few examples are provided. Following is a statement from:

> A CIRCE graduate assistant . . .
>
> ". . . thank you for the opportunity to be there and act and react as a professional. I really don't know whether it was AERC or just the experience of working in a different way with the people on our own staff, but I feel much closer to CIRCE know."

> A University of Illinois faculty member who was a workshop participant . . .
>
> "Because of my involvement in the 1970 Adult Education Research Conference, I can now point with pride that I know much more about the work of CIRCE. You and your coworkers did an excellent job of representing the University of Illinois. I hope it may be my priviledge of working more closely with you in the future."

> A participant's report about the workshop . . .
>
> "Perhaps the most striking quality of the spirit of the workshop was the feeling that there was not the usual dichotomy of 'speaker and audience' but rather a gathering of 'colleagues and peer group.' All attending were interested in learning about the field, and the psychological atmosphere of the group was one which is as conducive to learning through active participation."

## C. Side Effects and Bonuses

Numerous unexpected and positive side effects were associated with the workshop, for both participants and staff. Many of the participant outcomes presented in the previous section were side effects, as viewed by the participants, because participant expectations for the workshop were relatively low due to little pre-workshop publicity. Thus participants gained unanticipated tangible, cognitive, affective, and collegial outcomes.

For the staff there were numerous side effects in conducting the workshop. First, the instructional materials developed for the workshop (e.g., notebook, case studies) have been used subsequently in a variety of ways. Second, the participation of a large percentage of CIRCE staff in a common task was unprecedented. Third, graduate student staff were provided a means (financially) for also attending the American Educational Research Association Convention, which was held concurrently in the same city. Fourth, data from professional educators involved in evaluation activities were obtained on the CIRCE Attitude Scale. Fifth, professional contacts were made or continued. Last (and least), an important side effect for the writers of this report occurred when they were guests of the Holiday Inn Central in the penthouse suite for one night.

## D. Costs

For every educational venture costs are incurred (i.e., incapacities are created, balance sacrificed, concepts slighted). The major cost to the participants of the workshop was the limitation of the content to one evaluation approach; that is, participants were extensively exposed to the Countenance Model of Evaluation (á là Stake), but not to competing models such as those of Tyler, Stufflebeam, and Taba.

For the staff (especially the director) a major cost was time and energy expended as a result of the uncertainty of USOE financial support. Every aspect of the workshop had to have a "back-up" plan in case USOE funding was not forthcoming.

For both participants and staff there were additional costs borne by their institutions and their clients during the workshop: Classes were not conducted, correspondence stacked up.

## IV. CONGRUENCIES

This section is designed to communicate the congruence between the intended and the observed aspects of the evaluation workshop. To facilitate this communication this section will specify the congruencies between the intended and observed antecedents (prior conditions), transactions (activities), and outcomes (results) of the workshop.

### A. Antecedents

#### 1. The Setting and Facilities

The evaluation workshop was held at the location and at the time intended. It was held at the Holiday Inn Central, Minneapolis, Minnesota. It began on the later part of the second day (February 28) of the Adult Education Research Conference and extended through March 2. Meeting rooms were those intended (see Section II, Part E of this report). However, the quality of all the facilities did not meet expectations. Hotel rooms and eating facilities were viewed as quite satisfactory by participants, but meeting rooms were considered by most to be only average.

#### 2. Objectives

In conceptualizing the possible purposes of the workshop, the workshop planners intended that the *primary objective* of the workshop was to broaden the participants' conceptual framework from which they approach problems of evaluation in adult education. This objective was reported to have been achieved "very well" by 17 percent of the participants, "quite well" by 68 percent of the participants, and "somewhat" by 15 percent. No participants indicated that this intended objective was "hardly" or "not at all" achieved.

Participants attributed the successful accomplishment of this general objective primarily to the quality of the instructional staff, the organization of the workshop, the content of the workshop, and the materials (e.g., readings, bibliography) distributed to the participants.

In addition to the above general workshop objective, there were several intended specific *instructional objectives* of the workshop. Table 2 presents these objectives and indicates the extent to which participants felt they were achieved.

TABLE 2

Participant Responses to Achievement of Instructional Objectives

| Intended Objective | Extent of Achievement | | | | |
|---|---|---|---|---|---|
| | Highly | Quite | Somewhat | Hardly | Not at all |
| To examine in detail the Stake Countenance Model of evaluation | 24% | 56% | 15% | 5% | --- |
| To practice using the Stake Model for the identification and categorization of variables. | --- | 28% | 51% | 21% | --- |
| To design evaluation plans for typical adult education programs. | 5% | 18% | 41% | 36% | --- |
| To distinguish between summative and formative evaluation procedures. | 15% | 55% | 25% | 3% | 2% |
| To compare and contrast research and evaluative styles of inquiry. | 23% | 53% | 20% | 4% | --- |
| *To ascertain the role and importance of communications in evaluation. | 24% | 29% | 34% | 13% | --- |

* Introduced as an objective after the original workshop proposal was written, but before the workshop was conducted.

These data indicate that participants generally felt that the intended instructional objectives were achieved. There is evidence to support the conclusion that the objectives of a conceptual nature were more fully accomplished than applicational objectives. This was to be expected because the emphasis of the workshop was the conceptual aspects of evaluation.

In recognition that participants might have *personal objectives* for participating in the workshop, evidence was obtained to ascertain the extent to which their personal workshop objectives were achieved. Thirteen percent of the participants reported that their personal objectives were "extremely" well achieved. Sixty-five percent indicated that their personal objectives were "quite" well achieved. Twenty-two percent indicated that their personal objectives were "somewhat" achieved. No participants reported their personal objectives to be "hardly" or "not at all" achieved.

In summary, the congruence between the intended objectives (general, instructional, and personal) and the extent to which they were attained, based on participant reports, warrants the conclusion that the intended objectives actually were met.

### 3. Participants

The types of participants expected to attend the workshop did attend, but not in the numbers expected. That is, the participants had the expected level of education, type and extent of work responsibility, and diversity of interests in the field of adult education; but only one-half of the expected number of participants attended the workshop.

Participants selected for the workshop were professional adult educators who had major research and evaluation responsibilities. It was expected that about 50 percent would be spending at least one-fourth of their time on research and evaluation. In actuality 52 percent were spending at least one-fourth of their time on research and evaluation activities. Furthermore, 30 percent reported spending at least one-half of their time on research and evaluation.

It was also expected that one-half of the participants would have a doctorate with most of the remainder having a masters degree. In actuality, 57 percent of the participants had a doctorate, while 39 percent had masters degrees and four percent reported the baccalaureate degree as the highest degree attained.

The median year in which participants received their highest degree was 1967. (See Appendix H for a copy of the instrument used to collect these and related data.) Sixty-three percent of the participants received their highest degree in adult education, 12 percent in other areas of education (e.g., administration, educational psychology), and the remainder in areas such as communications, sociology, and agriculture.

As anticipated, participants entered the workshop with general background knowledge of evaluation, but with little knowledge of recent writings in the area. Participants were familiar with the writings of Tyler, but not with those of Popham, Scriven, Stufflebeam, or Stake. As one would expect, the topic of educational objectives was highly familiar to participants, but topics such as summative evaluation, formative evaluation, evaluation models, and unobtrusive measures were not familiar to participants.

Approximately 150 persons were expected to attend the evaluation workshop part of the Adult Education Research Conference; only about 75 did.[3] Several factors explain this incongruence. First, the workshop planners could not publicize the workshop to attract participants, because the decision by USOE to fund the workshop was not made prior to the workshop; in fact, funding approval was granted about two weeks after the workshop had occurred. Thus, the ability for potential participants to ask for local funds to attend a USOE-sponsored workshop was not realized. A second factor was that potential participants expressed difficulty in obtaining travel funds from their local institutions. This was especially observed for state education agencies who do not have a regular travel allowance. A final explanation for the reduced number of participants could be that the Conference was not held in a city with an adult education department associated with the local university. In previous Conferences the students associated with such departments comprised a major number of Conference attendees.

## B. Transactions

The institutional activities intended for the workshop were for the most part achieved. A comparison of the actual workshop schedule (Appendix J) with the tentative outline of the workshop activities (Appendix A) indicates topical congruence with only small changes in format. That is, the intended topics were presented, but there were more joint sessions held with all participants than had been planned. This adjustment was made, in part, because fewer participants attended the workshop than had been expected. Also, the intended assignment of participants to instructional groups based on characteristics such as field of adult education, background in evaluation, level of education, and attitude toward evaluation did not occur because the necessary information could not be gathered prior to the workshop due to lack of funds. For the small sessions the planned use of Instructional Teams was utilized. One person, Miss Mary Anne Bunda, was added to the instructional staff. (See Appendix K for a list of the acutal instructional staff.) Miss Bunda's role was primarily evaluative in nature. That is, she observed sessions (see Appendix L for the instrument) and provided feedback to the instructional staff, especially feedback of a formative nature.

---

[3] Appendix I contains a roster of participants who registered for the Conference. Approximately 20 of these persons did not stay on for the evaluation workshop.

The instructional activities intended to take place prior to the workshop did not occur. Again, this was due to the lack of funding available prior to the workshop. Thus the plan to send a copy of Stake's Countenance Model, "The Countenance of Educational Evaluation," to all participants prior to the workshop, to prepare a summary of the Stake Countenance Model, and to prepare an "Evaluation Vocabulary List" for distribution directly before the workshop was not realized. Neither were the professionally-developed tapes on selected evaluation topics intended for use during the workshop developed (e.g., "Outline of Countenance Model," "Distinction Between Research and Evaluation," and "Formative and Summative Evaluation").

To compensate for the lack of planned prior workshop activities, an Evaluation Notebook was prepared. This Notebook contained references to evaluation articles, a copy of Stake's "The Countenance of Educational Evaluation," a history of evaluation, and a number of instructional handouts. The reported usefulness of this Notebook, both during and after the workshop, more than compensated for the lack of the intended pre-workshop activities.

A final intended instructional activity that did not occur was the video-tape depiction of a meeting between a program evaluation consultant and an adult education program director. Due to the cost of equipment rental for displaying such a tape, the activity was deemed inefficient. Instead, a case study was developed (No. 6) based on an actual interview between an evaluator and an adult education program director. This case study was substituted for the intended video-tape.

## C. Outcomes

For the most part, the outcomes presented in Section III of this report were intended, although the extent of the desirable outcomes was not expected. The instructional staff anticipated that the participants would gain new knowledge, that they would experience some attitude change, and that they would profit from the collegial interaction. But we did not anticipate overwhelming success in all areas. Much of the success, no doubt, should be attributed to the sophistication of the learners, an input not present in many previous workshop experiences.

Also, the participants apparently felt the quality of instruction was better than they had anticipated. Participants rated approximately equally the instructional staff as "very" and "quite" *enthusiastic*. No one rated the staff as being "somewhat," "hardly," or "not at all" enthusiastic. The participants rated the staff similarly with respect to being *prepared* and being *helpful and friendly*. In general, 67 percent of the participants rated the staff as "superior," 30 percent as "good" and three percent as "average."

The most surprising outcome obtained was related to the Evaluation Notebook. This resource was expected by the staff to be useful for their own future needs, but the extensive satisfaction the participants expressed toward the Notebook was indeed an unexpected outcome.

# V. JUDGMENTS OF WORTH

## A. Value of Outcomes

There are many persons from whom judgments about the value of the workshop outcomes could be elicited. They include the Director of CIRCE, the instructional staff, the participants, non-participants, and leaders in the field of adult education.

The Director of CIRCE and the instructional staff valued highly the participant outcomes, especially the attitude changes exhibited. An additionally significant outcome was the side effect of CIRCE staff working on a common task.

Participants (some of whom were leaders in the field) also valued highly the outcomes of the workshop. Seventy-three percent of the participants rated the AERC Evaluation Workshop substantially better than similar workshops they had previously *attended*. Of the participants who had *conducted* workshops (79%), approximately one-third indicated that the evaluation workshop was substantially better than workshops they had conducted themselves. The remainder indicated it was about the same. In general, 29 percent of the participants felt the workshop had "much" impact on the field; 63 percent indicated "some" impact; and eight percent indicated "little" impact. No participants indicated that the workshop had either "great" or "no" impact--an important reality indicator.

It is also important to note that six letters were received by the director of the workshop from persons who had not participated. Each writer indicated that he regretted not having been in attendance.

## B. Relevance of Objectives to Needs

The relevance of the workshop objectives to needs can be viewed differently by different spokesmen. For the Adult Education Association, evaluation to improve program effectiveness is one of the major concerns. To a leader in the field, "The AERC Evaluation Workshop represents a significant attempt to meet an important need in the field." To approximately two-thirds of the participants attending the workshop the importance of AERC conducting a training workshop as part of its annual Conference was viewed as "extremely important." Not one participant indicated that this activity was not at least "somewhat important." Finally, the relevance of the objectives of the workshop were also of significant importance to CIRCE needs. The need for a continuing dialogue between educational practitioners (adult educators) and evaluation theoreticians (CIRCE) is desirable.

## C. Relation to Standards

It was anticipated that the worth of the workshop would be judged against many implicit standards held by different individuals. An attempt was made to make explicit some of these standards against which the workshop would be judged. They included the USOE, Research Training Branch guidelines governing proposals; the statements by leaders in the field of adult education; and the number of regional representation of participants at the workshop.

The funding of the workshop by USOE, Research Training Branch, warrants the conclusion that the workshop proposal met their standards. Also the statements by leaders in the field regarding the merit of the workshop (discussed earlier) implies some tacit approval of it. The participation standards of number and regional representation were not met as was desired. Most of the reasons for not meeting these standards (especially number) were examined earlier, but these deficiencies were in large part attributed to financial constraints. The lack of increased regional representation is also attributed to lack of pre-workshop publicity because of uncertain funding.

## D. Usefulness of Evaluation Information Gathering

As is now evident to the reader of this report, the evaluation information gathered was extensive. The extent to which it is useful to the various audiences for whom this report is written can only be estimated. Much depends on the expectations of these audiences. The fact that approximately 20 requests have been made for this report by adult educators in the field (other than workshop participants) implies that this report might be of use to this audience. To be sure the director of the workshop, and the staff, regard highly the usefulness of the data gathered. Gathering extensive data results in some negative side effects, however, the report was not completed on time.

# APPENDIX A: Tentative Outline of Workshop Activities

## Tentative Outline of Workshop Activities

### First Day

#### Morning

| | |
|---|---|
| 9:30 - 10:00 | Opening Coffee |
| 10:00 - 10:15 | Staff introduction, material distribution, and workshop plan and procedures |
| 10:15 - 10:45 | Lecture: Rationale and History of Evaluation |
| 10:45 - 11:15 | Lecture: Distinguishing Between Research and Evaluation |
| 11:15 - 12:05 | Lecture: Summative and Formative Evaluation |
| 12:05 - 12:15 | Evaluation No. 1 (Morning session) |

#### Afternoon

| | | | |
|---|---|---|---|
| 12:15 - 1:30 | Lunch | | |
| 1:30 - 2:15 | Group Participation: Discussion of CIRCE Attitude Scale Profiles | | |
| 2:15 - 2:30 | Lecture: Role of Evaluation Models | | |
| 2:30 - 3:15 | Lecture: Stake's Countenance Model of Evaluation | | |
| 3:15 - 3:30 | Coffee--Evaluation No. 2 (Unobtrusive listening and asking about Countenance lecture) | | |
| 3:30 - 4:00 | Lecture: Continuation of Stake's Countenance Model | | |
| | Group No. 1 | Group No. 2 | Group No. 3 |
| 4:00 - 4:30 | Objectives | Scaling | Field Methods |
| 4:30 - 5:00 | Communications | Unobtrusives | Scaling |
| 5:00 - 5:10 | Evaluation No. 3 (Afternoon Session) | | |
| 5:10 - 7:30 | Dinner | | |

#### Evening

| | |
|---|---|
| 7:30 - 9:30 | Video-taped adult education evaluation problem with variable identification (a la Stake). One evaluation episode divided into three 25-minute-video-presentations each followed by 15 minutes of discussion.<br>Distribution of exemplar-prototype solution. |

9:30 - Individual consultation and discussion.

Second day

Morning

| | Group No. 1 | Group No. 2 | Group No. 3 |
|---|---|---|---|
| 8:30 - 9:00 | Scaling | Objectives | Unobtrusives |
| 9:00 - 9:30 | Field | Communications | Objectives |
| 9:30 -10:00 | Unobtrusives | Field Methods | Communications |

| | |
|---|---|
| 10:00 -10:20 | Coffee |
| 10:20 -11:30 | Small Group: Application of Stake Countenance Model to 5 adult education evaluation case problems (Evaluation No. 4). Individuals to be assigned to form maximum diversity with respect to CIRCE Attitude Scale profile. |
| 11:30 -12:00 | Group Discussion: Common case discussion across three Instructional Groups with 5 groups of 30 participants each. |

Afternoon

| | |
|---|---|
| 12:00 - 1:00 | Lunch (Evaluation No. 5, Ascertain areas of evaluation to be covered). |
| 1:00 - 1:30 | Summary Discussion of Cases by Staff |
| 1:30 - 3:00 | Open for discussion of similar and/or related topics as needed, i.e., shown by formative evaluation of the workshop. |
| 3:00 - 3:20 | Coffee |
| 3:20 - 4:00 | Closing comments including second administration of CIRCE Attitude Scale (Evaluation No. 6) |
| 4:00 | End of Workshop |

# APPENDIX B: CIRCE Attitude Scale

**CIRCE Attitude Scale 1.4a** **Name**____________________________________

Different people have different ideas about the evaluation of educational programs. Some believe that maintaining a good school and improving instruction require carefully planned evaluation. Others believe that evaluation activities interfere with teaching and learning, doing more harm than good.

Different people see different purposes for educational evaluation. Certain people are oriented more to pupil behaviors or to classroom conditions or to other aspects of the program.

Responses to the items on this attitude scale provide us with 6 scale scores. When plotted on the profile sheet below they are expected to indicate the respondent's attitudes toward educational evaluation.

**Directions for Self Scoring**

Start in the opposite corner of this page. For each scale check your sheet to see how you responded to each of the eleven items. For Example, with SCALE V how did you mark Item #2? If you marked it "A" put a check in the parentheses. Put the number of checks in the box. Mark each horizontal scale (at the right) at the number-point shown in its box. Draw your profile by connecting your scores on the five scales, I-V. Then find your CONFIDENCE score.

**SCALE I**

3 A ( )
4 D ( )
11 D ( )
16 A ( )

22 D ( )
26 A ( )
27 D ( )
30 D ( )

31 A ( )
32 D ( )
39 D ( )

Total ☐

**I. A RESEARCH** orientation to Evaluation 0 1 2 3 4 5 6 7 8 9 10 11

The person high on this scale appears to believe that evaluation should rely on precise measurement and statistical analysis to gain **general understanding** of why programs do or do not succeed.

**SCALE II**

1 A ( )
4 A ( )
6 D ( )
13 D ( )

15 A ( )
16 D ( )
22 A ( )
28 D ( )

30 A ( )
32 A ( )
35 A ( )

Total ☐

**II. A SERVICE** orientation to Evaluation 0 1 2 3 4 5 6 7 8 9 10 11

The person high on this scale appears to believe that evaluation should be designed according to the **needs of the educators** involved so as to aid them in their present work and future decisions.

**SCALE III**

5 A ( )
9 A ( )
17 A ( )
20 D ( )

22 A ( )
23 A ( )
24 A ( )
34 A ( )

36 D ( )
37 A ( )
42 A ( )

Total ☐

**III. A TEACHING** orientation to Evaluation 0 1 2 3 4 5 6 7 8 9 10 11

The person high on this scale appears to believe that evaluation should be focused considerably on the **quality of teaching** and should discover the intrinsic merit in facilities and in instruction.

**SCALE IV**

7 D ( )
9 D ( )
10 A ( )
14 A ( )

15 D ( )
17 D ( )
19 A ( )
36 A ( )

37 D ( )
41 A ( )
43 A ( )

Total ☐

**IV. OBJECTIVES** orientation to Evaluation 0 1 2 3 4 5 6 7 8 9 10 11

The person high on this scale appears to believe that instruction, and therefore evaluation, should be focused considerably on apriori statements of objectives, that the merit of the program is largely indicated by the **success of students** in reaching those objectives.

**SCALE V**

Item 2 A ( )
4 A ( )
6 A ( )
12 D ( )

21 A ( )
27 A ( )
28 D ( )
31 D ( )

34 A ( )
39 A ( )
44 A ( )

Total ☐

**V. A JUDGMENT** orientation to Evaluation 0 1 2 3 4 5 6 7 8 9 10 11

The person high on this scale appears to believe that educational evaluation is largely a matter of establishing the worth of the program for various purposes **as perceived** by various groups of persons in and around the program.

**START** →→→→→

To obtain an overall....................**CONFIDENCE IN EVALUATION score,**
do the same thing with
the check-list at the right.

0 1 2 3 4 5 6 7 8 9 10 11

☐

**Confidence Scale**

Item 8 D ( )
12 D ( )
15 A ( )
18 D ( )

29 A ( )
34 A ( )
40 D ( )
45 A ( )

46 A ( )
47 A ( )
48 A ( )

**CIRCE Attitude Scale No. 1.4** Name________________________

**Attitudes toward Educational Evaluation.** Below are a number of statements about the evaluation of educational programs. A program can be a lesson, a course, a whole curriculum, or any training activity. Consider each statement as a statement of opinion. If you agree at least a little bit with the statement, circle the letter A. If you disagree even a little bit with the statement, circle the letter D. If you both agree and disagree, or if you have no opinion, leave the letters uncircled.

**A = AGREE** **D = DISAGREE** **Blank = Neither**

1. A D The **major** purpose of an educational evaluation study should be to gather information that will be helpful to the educators.
2. A D It is important for the program evaluator to find out how well various people like the program.
3. A D Generally speaking, an educational program should be evaluated with reference to one or more "control" programs.
4. A D The evaluator should accept the responsibility of finding the strongest, most defensible, and publicly attractive points of the program.
5. A D In evaluating a program, it is at least as important to study and report on the types of teaching as it is to study and report on the amount of learning.
6. A D The evaluator should draw a conclusion as to whether or not the goals of the program are worthwhile.
7. A D It is more important to evaluate a program in comparison to what other programs do than to evaluate it with reference to what its objectives say it should do.
8. A D Principals and superintendents should not gather data about the quality of instruction in the classroom.
9. A D The task of putting educational objectives into writing is **more** the responsibility of the evaluator than that of the educator.
10. A D It is essential that the full array of educational objectives be stated before the program begins.
11. A D Evaluation studies would improve if they gathered more kinds of information, even if at the expense of gathering less reliable information.
12. A D Evaluators should ignore data that cannot be objectively verified.
13. A D Education should have more of an engineering orientation than it now has.
14. A D The job of an evaluator is mostly one of finding out how well students learn what they are supposed to learn.
15. A D Evaluation should aid an educator in revising his goals even while the program is in progress.
16. A D The process of decision-making about the curriculum is one of the weakest links in the present operation of the schools.
17. A D Educators have some important aims that **cannot** be stated adequately by anyone in terms of student behaviors.
18. A D Information from an evaluation study is not worth the trouble it makes.
19. A D The first job in instruction is the formulation of a statement of objectives.
20. A D A teacher should tell his students any and all of his teaching objectives.
21. A D The **major** purpose of educational evaluation is to find out the worth of what is happening.
22. A D The evaluator should be a facilitator more than a critic or reformer or scholar.
23. A D Some school experiences are desirable because they round out a child's life—whether or not they increase his competence or change his attitudes.

Please turn over

24. A D An evaluator should find out if the teaching is in fact the kind that the school faculty expects it to be.

25. A D Whether or not an evaluation report is any good should be decided pretty much on the same grounds that research journal editors use to decide whether or not a manuscript should be published.

26. A D The **main** purpose of evaluation is to gain understanding of the causes of good instruction.

27. A D Description and value judgment are equally important components of evaluation.

28. A D In conducting an evaluation, there is no justification for the exercise of subjective judgment of any kind by the evaluator.

29. A D Educational evaluation is a necessary step in the everyday operation of the school.

30. A D The strategy of evaluation should be chosen primarily in terms of the particular needs the sponsors have for evaluation data.

31. A D The educational evaluator should attempt to conceal all of his personal judgment of the worth of the program he is evaluating.

32. A D The sponsor of an evaluation should have the final say-so in choosing or eliminating variables to be studied.

33. A D The main purpose of educational evaluation is to find out what methods of instruction work for different learning situations.

34. A D Parents' attitudes should be measured as part of the evaluation of school programs.

35. A D An evaluator finds it almost impossible to do his job without intruding upon the operation of the program at least a little.

36. A D All important educational aims can be expressed in terms of student behaviors.

37. A D Some educational goals are best expressed in terms of teacher behaviors.

38. A D It is essential that evaluation studies be designed so that the findings are generalizable to other curricula.

39. A D An evaluation study should pay less attention to the statistical significance of a finding than an instructional research study would.

40. A D Evaluation interferes with the running of schools more than it helps.

41. A D Little evaluation planning can be done before you get a statement of instructional objectives.

42. A D The leader of an evaluation team should be a teacher.

43. A D The entire school day and the entire school experience should be divided up and assigned to the pursuit of stated educational goals.

44. A D An evaluation of an educational program should **include** a critical analysis of the value of the goals of the program.

45. A D Every teacher should have formal ways of gathering information about the strengths and shortcomings of his instructional program.

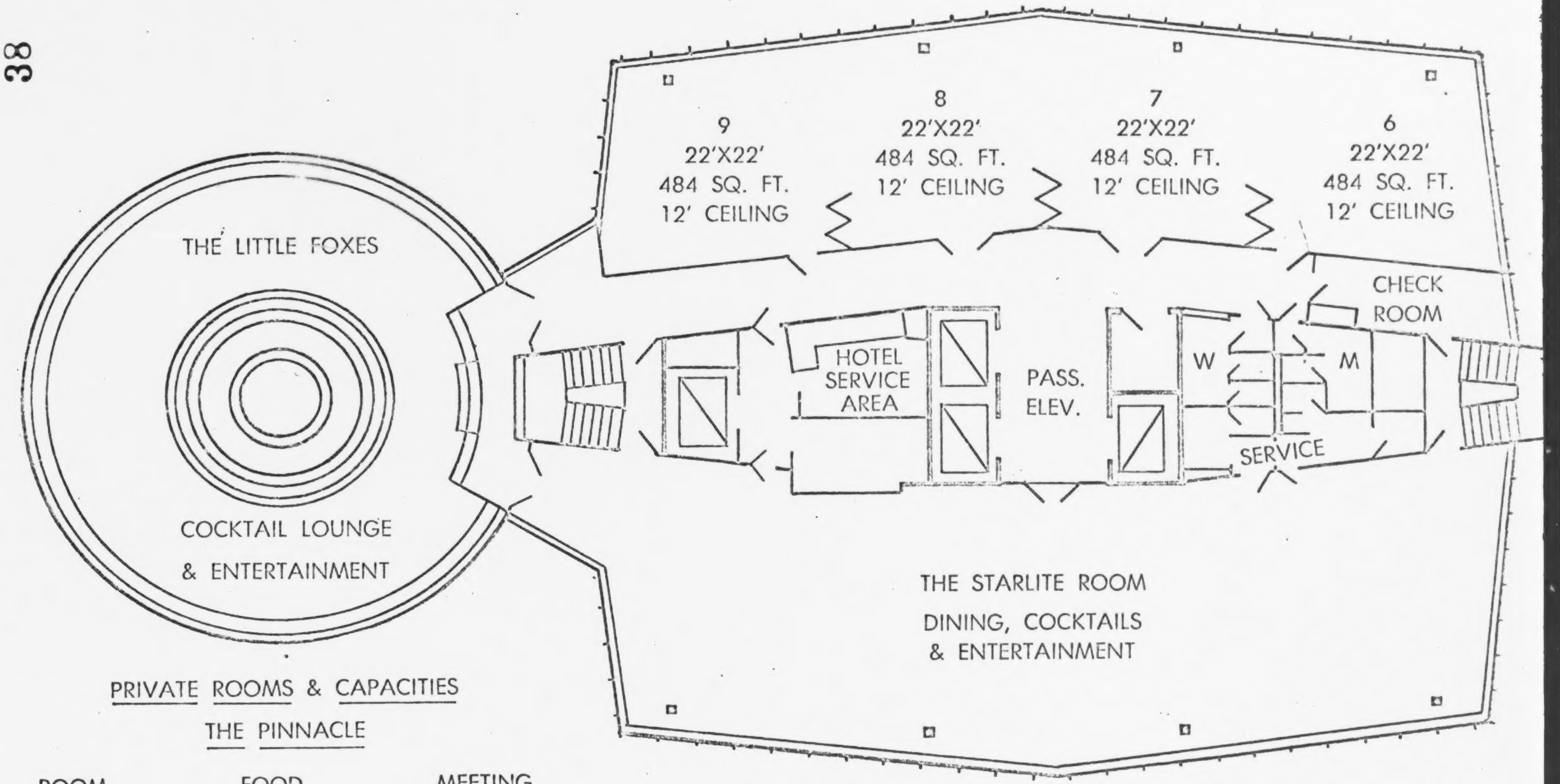
46. A D Money spent on evaluation contributes more to the improvement of education than any other expenditure.

47. A D There just is no way that careful and honest evaluation can hurt a school program.

48. A D If an evaluation study is well designed, the primary findings are likely to improve decisions made by administrators, teachers, and students themselves.

49. A D When the evaluator has to choose between helping this staff run its program better and helping educators everywhere understand all programs a little better he should choose the latter.

## APPENDIX C: Floor Plan of Evaluation Workshop Meeting Area


9
22'X22'
484 SQ. FT.
12' CEILING
8
22'X22'
484 SQ. FT.
12' CEILING
7
22'X22'
484 SQ. FT.
12' CEILING
6
22'X22'
484 SQ. FT.
12' CEILING
THE LITTLE FOXES
COCKTAIL LOUNGE
& ENTERTAINMENT
HOTEL
SERVICE
AREA
PASS.
ELEV.
W
M
CHECK
ROOM
SERVICE
THE STARLITE ROOM
DINING, COCKTAILS
& ENTERTAINMENT

PRIVATE ROOMS & CAPACITIES

THE PINNACLE

| ROOM | FOOD | MEETING |
|---|---|---|
| NO. 6 | 40 | 75 |
| NO. 7 | 40 | 75 |
| NO. 8 | 40 | 75 |
| NO. 9 | 40 | 75 |

BANQUET STYLE – 240

**14TH FLOOR – THE PINNACLE**
**HOLIDAY INN OF MINNEAPOLIS-CENTRAL**
**MINNEAPOLIS, MINNESOTA**

## APPENDIX D: AERC Evaluation Workshop Notebook

# AERC EVALUATION WORKSHOP NOTEBOOK
## Table of Contents

Content | Page
---|---
Selected References on Evaluation . . . . . . . . . . . . . . . . . . . . | 1
Prototypes of Curriculum Evaluation . . . . . . . . . . . . . . . . . . . | 8
The Growth of Evaluation Methodology. . . . . . . . . . . . . . . . . . . | 9
The Countenance of Educational Evaluation . . . . . . . . . . . . . . . . | 43
Generalizability of Program Evaluation: The Need for Limits. . . . . | 61
Educational Objectives. . . . . . . . . . . . . . . . . . . . . . . . . . | 63
Multiple Criterion Measures for Evaluation of School Programs . . . . | 64
Format for an Evaluation Report for an Educational Program. . . . . . | 70
Evaluation Reports. . . . . . . . . . . . . . . . . . . . . . . . . . . . | 71
AERA Monograph Series on Curriculum Evaluation. . . . . . . . . . . . | 78
EPIE Brochure . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . | 79
Map of Meeting Rooms. . . . . . . . . . . . . . . . . . . . . . . . . . . | 80

# AERC EVALUATION WORKSHOP

## Selected References on Evaluation

### History of Evaluation

**Glass, G. V. The Growth of Evaluation Methodology. AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNally, 1971 (in press).

*Merwin, J. C. Historical Review of Changing Concepts of Evaluation. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969. Pp. 6-25.

### Rationale and Philosophy of Evaluation

Cohen, D. K. Politics and Research: The Evaluation of Large Scale Social Action Programs in Education. Review of Educational Research, 1970, (in press).

*Cronbach, L. J. Course Improvement through Evaluation. Teachers College Record, 1963, 64, 672-683.

*Gooler, D. D. and Grotelueschen, A. Process Accountability in Curriculum Development. (Special Issue No. 1 on Curriculum Development), Curriculum Theory Network. Toronto: OISE, 1970 (in press).

*Hastings, J. T. Curriculum Evaluation: the Whys of the Outcomes. Journal of Educational Measurement, 1966, 3, 27-32.

*Scriven, M. The Methodology of Evaluation. AERA Monograph Series on Curriculum Evaluation, Vol. 1. Chicago: Rand McNally, 1967, Pp. 39-83.

*Stake, R. E. Toward a Technology for the Evaluation of Educational Programs. AERA Monograph Series on Curriculum Evaluation, Vol. 1. Chicago: Rand McNally, 1967. Pp. 1-12.

*Stake, R. E. Objectives, Priorities, and Other Judgment Data. Review of Educational Research, 1970, (in press).

*Stake, R. E. and Denny, T. Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969. Pp. 370-390.

*In AERC Evaluation Workshop Library
**In AERC Evaluation Workshop Notebook

Taba, Hilda. The Functions of Evaluation. Childhood Education, 1939, 15, 245-251.

*Tyler, R. W. The Purposes of Assessment. In W. H. Beatty (Ed.), Improving Educational Assessment and An Inventory of Measures of Affective Behavior. Washington: Association for Supervision and Curriculum Development, NEA, 1969. Pp. 2-13.

## Evaluation Models

Alkin, M. C. Towards an Evaluation Model: A Systems Approach. Working Paper No. 4. Los Angeles: Center for the Study of Evaluation of Instructional Programs, University of California, 1967.

Guba, E. G. and Stufflebeam, D. L. Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action. Columbus: Evaluation Center (1712 Neil Avenue), Ohio State University, 1968.

National Study of Secondary School Evaluation. Evaluative Criteria, 1960 Edition. Washington: National Study of Secondary School Evaluation, 1960.

*Provus, M. Evaluation of Ongoing Programs in the Public School System. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969. Pp. 242-283.

**Stake, R. E. The Countenance of Educational Evaluation. Teachers College Record, 1967, 68, 523-540.

*Stufflebeam, D. L. Evaluation as Enlightenment for Decision Making. In W. H. Beatty (Ed.), Improving Educational Assessment and An Inventory of Measures of Affective Behavior. Washington: Association for Supervision and Curriculum Development, NEA, 1969. Pp. 41-73.

Taba, Hilda. Teaching Strategies and Cognitive Functioning in Elementary School Children. Cooperative Research Project No. 2404. San Francisco State College, San Francisco, 1966.

Taylor, A. P. and Maguire, T. O. A Theoretical Evaluation Model. The Manitoba Journal of Educational Research, 1966, 1, 12-17.

Tyler, R. W. General Statement on Evaluation. Journal of Educational Research, 1942, March, 492-501.

## Research vs. Evaluation

Cronbach, L. J. and Suppes, P. (Eds.) Research for Tomorrow's Schools: Disciplined Inquiry for Education. Toronto: MacMillan, 1969. Pp. 122-200.

Guba, E. G. Significant Differences. Educational Research Newsletter, American Educational Research Association. 1969, 20, No. 3, Pp. 4-5.

*Hemphill, J. K. The Relationships Between Research and Evaluation Studies. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969. Pp. 189-192.

Simon, G. B. What Kind of Evaluation? Educational Researcher Newsletter, American Educational Research Association. 1969, 20, No. 5, Pp. 9-10.

**Stake, R. E. Generalizability of Program Evaluation: The Need for Limits. Educational Product Report. 1969, 2, 39-40.

Stanley, J. C. Principles of Scientific Research Are Not Pertinent to Educational Evaluation? Educational Researcher Newsletter, American Educational Research Association. 1969, 20, No. 5, Pp. 8-9.

*Wardrop, J. L. Generalizability of Program Evaluation: The Danger of Limits. Educational Product Report. 1969, 2, 41-42.

## Summative and Formative Evaluation

*Bloom, B. S. Some Theoretical Issues Relating to Educational Evaluation. In R. W. Tyler (Ed.), Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969. Pp. 47-50.

*Scriven, M. The Methodology of Evaluation. AERA Monograph Series on Curriculum Evalution, Vol. 1. Chicago: Rand McNally, 1967. Pp. 40-43.

## Objectives in Evaluation

*AERA Monograph Series on Curriculum Evaluation. Vol. 3, Chicago: Rand McNally, 1969.

*Atkin, J. M. Some Evaluation Problems in a Course Content Improvement Project. Journal of Research in Science Teaching, 1963, 1, 129-132.

Broudy, H. S. Research and the Dogma of Behavioral Objectives. Paper read at the American Educational Research Association Meeting, Los Angeles, 1969.

Eisner, E. W. Educational Objectives: Help or Hindrance. The School Review, 1967, 75, 250-260.

Krathwohl, D. R. Stating Objectives Appropriately for Program, for Curriculum, and for Instructional Materials Development. Journal of Teacher Education, 1965, 16, 83-92.

Maguire, T. O. Value Components of Teachers' Judgments of Educational Objectives. AV Communication Review, 1968, 16, 63-86.

Maguire, T. O. Decisions and Curriculum Objectives: A Methodology for Evaluation. Alberta Journal of Educational Research, 1969, 15, 17-30.

## Communications

Gerbner, G. (Ed.) The Analysis of Communication Content: Developments in Scientific Theories and Computer Techniques. New York: Wiley, 1969.

*Stake, R. E. Language, Rationality, and Assessment. In W. H. Beatty (Ed.), Improving Educational Assessment and An Inventory of Measures of Affective Behavior. Washington: Association for Supervision and Curriculum Development, NEA, 1969. Pp. 14-40.

## Unobtrusive Measures and Observation

Campbell, D. T. Administrative Experimentation, Institutional Records, and Nonreactive Measures. In J. C. Stanley (Ed.), Improving Experimental Design and Statistical Analysis. Chicago: Rand McNally, 1967. Pp. 257-291.

Medley, D. M. and Mitzel, H. E. Measuring Classroom Behavior by Systematic Observation. In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963. Pp. 247-328.

Remmers, H. H. Rating Methods in Research on Teaching. In N. L. Gage (Ed.), Handbook of Research on Teaching. Chicago: Rand McNally, 1963. Pp. 329-378.

Sechrest, L. Use of Innocuous and Noninterventional Measures in Evaluation. Paper read at the American Educational Research Association Meeting, Chicago, 1968.

Simon, Anita and Boyer, E. G. Mirrors for Behavior: An Anthology of Classroom Observation Instruments. Philadelphia: Research for Better Schools, Inc., 1967.

*Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago: Rand McNally, 1966.

## Measurement in Evaluation

Metfessel, N. S. and Michael, W. B. A Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs. Educational and Psychological Measurement, 1967, 27, 931-943.

Sjogren, D. D. Measurement Strategies for Evaluation. *Review of Educational Research*, 1970. (in press).

*Stake, R. E. Objectives, Priorities, and Other Judgment Data. Review of Educational Research, 1970. (in press).

### Scaling

Guilford, J. P. *Psychometric Methods*. New York: McGraw-Hill, 1954.

Torgerson, W. S. *Theory and Methods of Scaling*. New York: Wiley, 1958.

### Measurement

Dowd, D. J. and West, Sarah C. An Inventory of Measures of Affective Behavior. In W. H. Beatty (Ed.), *Improving Educational Assessment and an Inventory of Measures of Affective Behavior*. Washington: Association for Supervision and Curriculum Development, NEA, 1969, Pp. 90-158.

Edwards, A. L. *Techniques of Attitude Scale Construction*. New York: Appleton-Century Crofts, 1957.

Lindquist, E. F. (Ed.) *Educational Measurement*. Washington: American Council on Education, 1951.

Nunnally, J. C. *Tests and Measurements*. New York: McGraw-Hill, 1959.

Shaw, M. E. and Wright, J. M. *Scales for the Measurement of Attitudes*. New York: McGraw-Hill, 1967.

Snider, J. G. and Osgood, C. E. (Eds.) *Semantic Differential Technique: A Sourcebook*. Chicago: Aldine, 1970. (in press)

Test Collection Bulletin. A Quarterly Digest of Information on Tests. Princeton, N. J.: Educational Testing Service.

### Survey Research

Festinger, L. and Katz, D. *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart and Winston, 1953.

Glock, C. Y. (Ed.) *Survey Research in the Social Sciences*. New York: Russell Sage Foundation, 1967.

Oppenheim, A. N. *Questionnaire Design and Attitude Measurement*. New York: Basic Books, 1966.

Payne, S. L. *The Art of Asking Questions*. Princeton, New Jersey: Princeton University Press, 1951.

Stephan, F. F. and McCarthy, P. J. Sampling Opinions, An Analysis of Survey Procedure. New York: Wiley, 1958.

### Evaluation Reports

*Gooler, D. D. Institute on Educational Program Evaluation: From Realization to Action. (An Account of an Institute.) Urbana, Ill.: CIRCE, University of Illinois, 1969.

*Hastings, J. T., Grotelueschen, A., and Sjogren, D. D. Aperiodic Report: Trainers of Teacher Trainers (TTT) Evaluation. Urbana, Ill.: CIRCE, University of Illinois, 1969.

*Lockerby, Florence (Ed.) The Report on the 1968 Summer Institute on Evaluation. Northfield, Illinois: Cooperative Educational Research Laboratory, Inc. (CERLI), 1968.

*Stake, R. E. An Opinion of the Evaluation of Title III. Report to the Committee of Special Consultants to the 1967-68 Title III Evaluation Study, Center for Instructional Research and Curriculum Evaluation, University of Illinois, Urbana, Ill, 1968.

*Stake, R. E., Glass, G. V, and Taylor, P. A. Report of the AERA 1966 Presession on Experimental Design. Urbana, Ill.: CIRCE, University of Illinois, 1966.

### Resource Materials

*AERA Monograph Series on Curriculum Evaluation. Chicago: Rand McNally.

*Beatty, W. H. (Ed.) Improving Educational Assessment and An Inventory of Measures of Affective Behavior. Washington: Association for Superivision and Curriculum Development, NEA, 1969.

Bloom, B. S. (Ed.) Taxonomy of Educational Objectives, Handbook I: Cognitive Domain. New York: Longmans, Green, 1956.

Buros, O. K. (Ed.) The Sixth Mental Measurements Yearbook. Highland Park, N. J.: Gryphon, 1965.

*Educational Product Report (EPR). Educational Products Information Exchange (EPIE Institute), 386 Park Avenue South, New York, N.Y. 10016.

Gage, N. E. (Ed.) Handbook of Research on Teaching. Chicago: Rand McNally, 1963.

Glass, G. V (Ed.) Review of Educational Research. Washington: American Educational Research Association.

Kerlinger, F. N. Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, 1964.

Krathwohl, D. R., Bloom, B. S., Masia, B. B. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay, 1964.

*Tyler, R. W. (Ed.) Educational Evaluation: New Roles, New Means. Sixty-eighth Yearbook of the National Society for the Study of Education, Part II. Chicago: University of Chicago Press, 1969.

# PROTOTYPES OF CURRICULUM EVALUATION

| PROTOTYPE | KEY EMPHASIS | PURPOSE | KEY ACTIVITIES | KEY VIEWPOINT USED TO DELIMIT STUDY | OUTSIDE EXPERTS NEEDED | EXPECTED TEACHING STAFF INVOLVEMENT | RISKS | PAYOFF |
|---|---|---|---|---|---|---|---|---|
| Ralph Tyler's Evaluation Model | Instructional objectives | To measure student progress toward objectives | Specify objectives; measure student competence | Curriculum supervisor; teacher | Objectives specifiers; measurement specialists | Conceptualize objectives; give tests | Oversimplify school aims; ignore processes | Ascertain student progress |
| School Accreditation Model | Staff self-study | To review content and procedures of instruction | Discuss program; make professional judgments | Classroom teacher; administrator | None, unless authentication by outside peers needed | Committee discussions | Exhaust staff; ignore values of outsiders | Increase staff leadership responsibility |
| Bob Stake's Countenance Model | Description and judgment data | To report the ways different people see the curriculum | Discover what audience wants to know about; observe; gather opinions | Audience of final report | Journalists; social psychologists | Keep logs; give opinions | Stir up value conflicts; ignore causes | Broad picture of curriculum and conflicting expectations |
| Dan Stufflebeam's CIPP Model | Decision-making | To facilitate rational and continuing decision-making | Identify upcoming alternatives; study implications; set up quality-control | Administrator; director | Operations analysts | Anticipate decisions, contingencies | Overvalue efficiency; undervalue student aims | Curriculum sensitive to feedback |
| Hilda Taba's Social Studies Evaluation Model | Cause and effect relationships | To seek simple but enduring explanation of what works | Exercise experimental control and systematic variation | Theorist; researcher | Research designer; statistical analysts | Tolerate experimental constraints | Artificiality; ignore personal values | Get rules for developing new programs |

REFERENCES:

TYLER, RALPH W. General Statement on Evaluation. *Journal of Educational Research*, March, 1942, 492-501.

NATIONAL STUDY OF SECONDARY SCHOOL EVALUATION. *Evaluative Criteria, 1960 Edition*, National Study of Secondary School Evaluation, Washington, D. C.

STAKE, ROBERT E. The Countenance of Educational Evaluation. *Teachers College Record*, 68, 1967, 523-540.

STUFFLEBEAM, DANIEL L. Evaluation as Enlightenment for Decision Making. In W. H. Beatty (Ed.), *Improving Educational Assessment and An Inventory of Measures of Affective Behavior*. Washington: Association for Supervision and Curriculum Development, NEA, 1969. Pp. 41-73.

TABA, HILDA. *Teaching Strategies and Cognitive Functioning in Elementary School Children.* Cooperative Research Project No. 2404. San Francisco State College, San Francisco, 1966.

Prepared by R. E. Stake, CIRCE, October 1969

# The Growth of Evaluation Methodology

Gene V Glass

*Laboratory of Educational Research*
*University of Colorado*

The biological principle of allometry is that the form of an organism limits its growth. It is a distinctive feature of living things that they stop growing at some point (unlike stalagmites and stalactites, for example). Imagine that the genetic code (the form) of an organism dictates that it grow cubically. If the environment of such an organism (availability of food, metabolic rate, etc.) permits it 8 units of size for its phenotype, that organism can only grow to 2 units of size alone each dimension. If an organism must grow spherically, 8 units of growth limit its diameter to about 2.5 units. If, however, the phenotype to the organism is destined to be square and one cell thick, its 8 units of phenotypic material permit it to cover an immense area at maturity.

An insect breathes through its "skin". This is a major factor limiting its size. If an insect were as large as a man, its oxygen assimilating surface could not support the insect's needs because in growing from an eighth of an inch to six feet, its volume would grow so much faster than its surface that it would suffocate. The convoluted human lung contains an oxygen assimilating surface so large that it can permit growth to nearly six feet. Thus, form limits growth in biology.

Kenneth Boulding (1953, pp. 21-32) relates the biological principle of allometry to a range of non-biological phenomena. The principle can be validly applied to the study of organizations. The form taken by a social organization limits its growth. The genotype of an organization is contained in such things as the technology available to it and its images of the future. An organization forced to rely on face-to-face transmission of information semi-weekly to all its members can probably not grow to more than 100 members. Use of the telephone might permit the organization to double in size. If the organization can exist with only annual communication between members, it may grow to huge proportions. Prior to 1860 the federal government never had more than 5,000 employees. The technology available for making organizations work in that day would not have permitted a much larger group of employees. The direction and rate of growth of organizations is also governed by their "self-concept." Organizations have an image of themselves in the present and the future. General Motors could quickly become the world's leading producer of women's lingerie, but it seems safe to wager that such a role is inconsistent with General Motors' self-concept and that they will continue to produce cars.

*Reproduced by permission of the author for distribution to participants of the Evaluation Workshop of the Adult Education Research Conference, Minneapolis, Minnesota, March, 1970.

Allometry governs the growth of organizations of persons, things and even ideas. The growth of a scholarly discipline is partially governed by the form it assumes. Its form is contained in a genetic code set by accident and design by those who launch the discipline. The genes in this genetic code determine such things as the phenomena of interest, methods and techniques used to study these phenomena, the scope of the discipline, etc.

The principle of allometry has an obvious extension in the social realm: form limits growth which limits utility. Some economic, social and scientific organizations assume a form that stunts their growth and limits their social utility. Other organizations grow misshapen and are wasted.

My purpose in this paper is to identify four models of educational evaluation, to determine their form (i.e., their technology and view of their future), and to judge their potential growth and social utility. I will examine the Tylerian model, the Accreditation model, the Management-Systems model, and the Composite-Goal model.

## Educational Research and Evaluation Distinguished

Before moving to the analysis of four evaluation models, it will be well to distinguish educational evaluation from the assortment of activities called "educational research". The attempt to distinguish research and evaluation is neither idleness nor pedantic Aristotelianism. It is clear that abstract, verbal definitions do influence behavior and that some educational research is poorly done because it is called "evaluation" but that far more evaluation activity is wasted because it is regarded as an educational research project.

Simple verbal definitions of research and evaluation are so non-exclusive of one another as to be worthless. Defining research as "the search for understanding of phenomena in systems of related phenomena where 'understanding' is defined as the ability to predict and control" is inadequate. Evaluation seeks the ability to predict and control as well, yet we still feel that it aspires to predict and control different things in different ways from the content and methods of research.

The difficulty in distinguishing between educational research and educational evaluation is that there exists so few examples of each type of activity in a pure form. Most large empirical studies on educational problems combine evaluative and research questions in varying proportions. An attempt to sort educational studies into two piles would yield the same confusing result that any similar attempt to so distinguish two concepts in the social sciences would produce: one small pile labeled research, one small pile labeled evaluation, and a large pile labeled other. The minor confusions that zoologists must face in fitting whales and porpoises into their categorizing schemes are faced in abundance by taxonomists in the social and behavioral sciences.

Though it is nearly hopeless to discover what research and evaluation are by studying projects or studies, individual problems or questions can be more successfully categorized as research or evaluation. However, even at this level the distinction is obscured by the fact that the two activities are only differentiable with respect to continuous characteristics (e.g., the inquirer's motives, the relationship of the findings to other knowledge, the use made of the findings) so that one activity fades imperceptibly into the other. Both research and evaluation are mixtures of empirical and rational inquiry; they make use of many of the same techniques (inferential statistical analysis, experimental design, psychometrics, survey analysis, etc.); both activities eventuate in findings that are useful and true to varying degrees. And yet, research and evaluation are distinctly different enterprises.

The authors of _Research for Tomorrow's Schools_: _Disciplined Inquiry for Education_ (1969, pp. 20-21) distinguished between _decision-oriented_ and _conclusion-oriented_ research:

> In a decision-oriented study the investigator is asked to provide information wanted by a decision-maker: a school administrator, a governmental policy-maker, the manager of a project to develop a new biology textbook, or the like. The decision-oriented study is a commissioned study. The decision-maker believes that he needs information to guide his actions and he poses the question to the investigator. The conclusion-oriented study, on the other hand, takes its direction from the investigator's commitments and hunches. The educational decision-maker can, at most, arouse the investigator's interest in a problem. The latter formulates his own question, usually a general one rather than a question about a particular institution. The aim is to conceptualize and understand the chosen phenomenon; a particular finding is only a means to that end. Therefore, he concentrates on persons and settings that he expects to be enlightening.

Conclusion-oriented inquiry is much like what is here referred to as research; decision-oriented inquiry typifies evaluation as well as any three words can.

Not altogether satisfactorily we can say that educational evaluation attempts to assess the _worth_ of a thing and education research attempts to assess the scientific _truth_ of a thing. Except that truth is highly valued and hence that which possesses it is worthy, this distinction serves fairly well to discriminate research and evaluation. The distinction can be made less ambiguous if "worth" is taken as synomous with "social utility" (which increases with increases in health, happiness, life expectancy, etc., and decreases with increases in privation, sickness, ignorance, etc.) and if "scientific truth" is identified with two of its many forms: 1) empirical verifiability of a general phenomenon* with accepted methods of inquiry;

---

*A "general phenomenon" is one that is evidenced or can be found in a wide range of ostensibly different settings and is invoked as the touchstone of a scientific concept. Without such a qualification, determining empirically that your keys are lost would be an "assessment of scientific truth." The concept of the generality of expected findings is important for distinguishing evaluation and research; it is of great practical importance in designing an evaluation study. (See Stake, 1969).

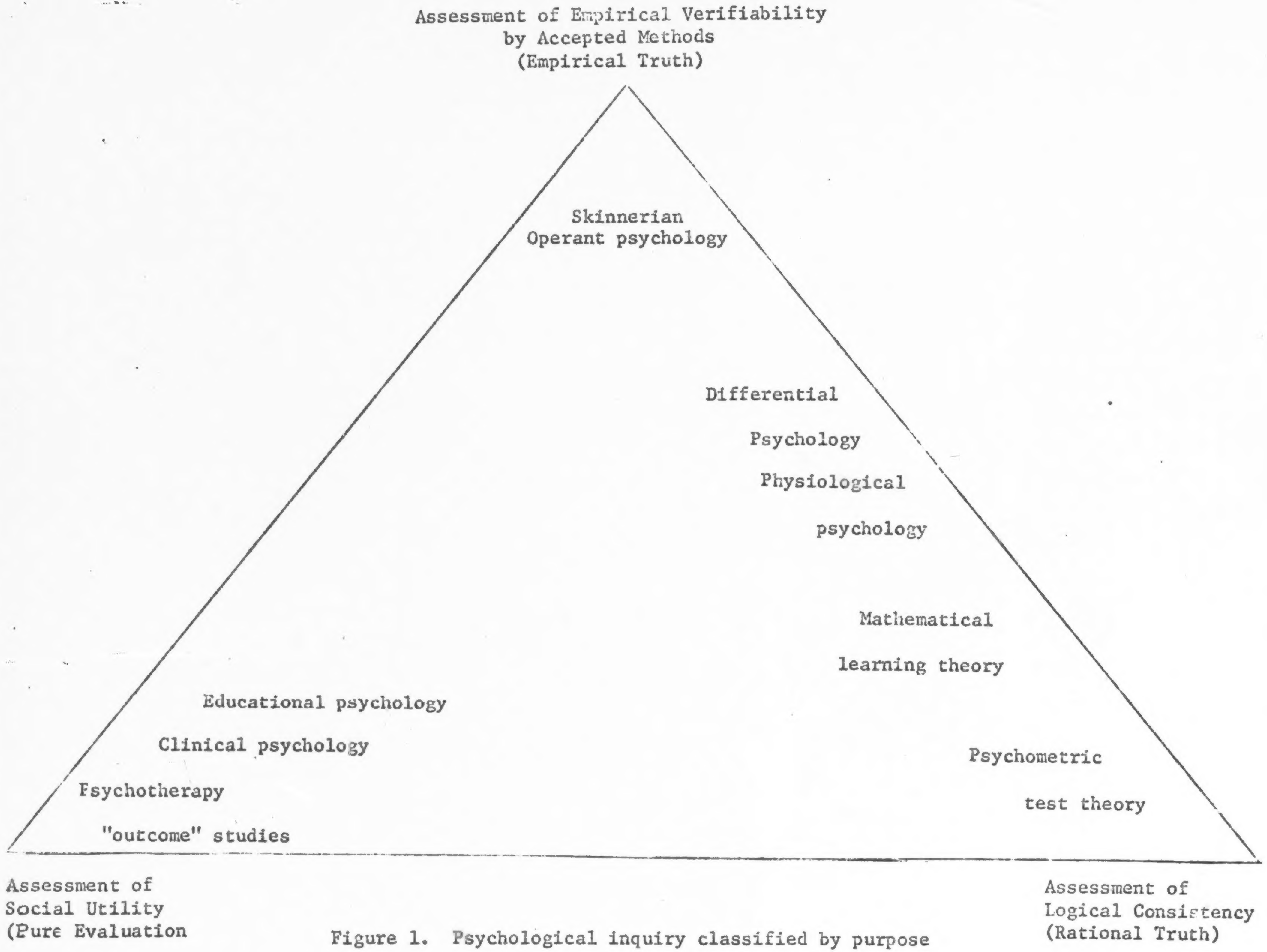

Assessment of Empirical Verifiability
by Accepted Methods
(Empirical Truth)
Skinnerian
Operant psychology
Differential
Psychology
Physiological
psychology
Mathematical
learning theory
Psychometric
test theory
Educational psychology
Clinical psychology
Fsychotherapy
"outcome" studies
Assessment of
Social Utility
(Pure Evaluation
Assessment of
Logical Consistency
(Rational Truth)

Figure 1. Psychological inquiry classified by purpose

2) logical consistency. The distinction between assessing worth (evaluation) and scientific truth (research) so defined now takes on more meaning.

Evaluation is that activity which seeks directly to assess social utility. Research may yield evidence of social utility, but only indirectly -- because empirical verifiability of general phenomena and rational consistency may eventually be of substantial social utility. A litmus test for discriminating an evaluator and a researcher is to ask whether the inquiry would be regarded as a failure if it produced no information or whether the phenomenon studied was useful or useless. A researcher answering qua research will probably say No.

In the above view, inquiry is seen as directed toward the assessment of three distinct properties of a phenomenon: 1) empirical verifiability of phenomena by accepted methods; 2) logical consistency; 3) social utility. Most disciplined inquiry aims to assess each property in varying degrees. In Figure 1, several areas of inquiry within psychology are classified with respect to the degree to which they seek to assess each of the above three phenomena. The distance of a point in the triangle from each of the angles of the triangle is inversely related to the extent to which the property represented by that angle is sought by the inquiry at that point.

## The Tylerian Model of Evaluation

The earliest model for curriculum evaluation emerged from the Eight-year Study (the Commission on the Relation of School and College). This model was developed during the 1930's by Ralph W. Tyler and the evaluation staff of the Eight-year Study. The methodology of evaluation developed by Tyler and his associates was presented in publications by Smith and Tyler (1942) and Tyler (1951). The following guidelines constitute the Tylerian Evaluation model:

1) Formulate objectives. Determine the broad goals of the program.
2) Classify objectives. Develop a typology of objectives so an economy of thought and action may be achieved.
3) Define objectives in behavioral terms. This feature has become the cornerstone of the Tyler model. "Modern" methodologies of evaluation which rest heavily upon the specific, behavioral statement of objectives have not moved beyond Tyler's thoughts on evaluation in the Eight-year Study.
4) Suggest situations in which achievement of objectives will be shown.
5) Develop or select appraisal techniques (standardized tests, ad hoc tests, questionnaires, etc.).
6) Gather and interpret performance data. The final step in the evaluation process involved the measurement of student performance and the comparison of performance data with behaviorally stated objectives. The program was presumably praised for its successes (so determined) and condemned for its failures.

Tylerian curriculum eveluation places almost exclusive priority on pupil behaviors. Objectives must be stated in behavioral terms, and test data on performance of the desired behaviors is all that deserves the evaluator's attention. The curriculum evaluators are proud that they evaluate the ends of instruction and not just the means toward those ends.

The claim of modern curriculum evaluators that they evaluate the ends of education (pupil behaviors) and not the mere means cannot be justified. With the possible exception of literacy (verbal and quantitative), most significant objectives of instruction are behaviors that will be manifested after (by years, perhaps) formal instruction has ceased. Some objectives are literally unobservable, e.g., "that a student should cast an intelligent, rational ballot in an actual secret election upon reaching his majority." The example is significant; the way a man conducts his private affairs is similarly not practically or ethically observable.

For a significantly large portion of the curriculum -- the majority of it perhaps -- the actual behaviors that educators strive to bring about cannot be observed. Therefore, instruction must be evaluated by observing proxy events or behaviors. (Proxy behaviors are those that take the place or "stand in for" the ultimate objectives which cannot economically or ethically be observed.) An example of a proxy event for rational voting in elections might be a student casting a ballot that is not secret, after he has heard the campaign speeches on Junior-high County Government Day.

Performance on a proxy event is only circumstantial evidence of the same performance on the actual or ultimate event. Much evaluation by means of assessing performance toward behaviorally stated objectives produces only circumstantial evidence that the student has or will attain the ultimate objective of instruction, which generally involves transfer or generalization to a non-school setting.

The consequence of accepting one sort of circumstantial evidence in evaluation is that other types of circumstantial evidence must also be accepted. The "other types" may not necessarily involve pupil behaviors. That a particular lesson is logically relevant, that a school schedule is free of needless disruptions, and that tests are used for punitive purposes are also circumstantial evidence that students are or are not attaining the objectives of instruction. Thus, there are compelling reasons not to rely exclusively on the use of measurement of pupil behaviors on behaviorally stated objectives in curriculum evaluation. One must consider a much broader range of evidence. Observations and judgments must be made of the curriculum materials themselves, teachers, organizational plans, etc. In many instances these latter sources of evidence should take precedence over pupil behaviors.

Traditional thinking on educational evaluation held that judgments are subjective, and hence, are not suitable material from which to build an evaluation. Judgments are undeniably subjective, but they can be gathered and reported with objectivity. Moreover, the subjectivity of value judgments makes them important as determiners of the success of a program. It is beside the point to observe that a principal's judgment is subjective, when that principal's judgment that a program has worthless objectives causes him to undercut the program by withdrawing his support. Indeed, judgments, attitudes, and satisfactions are subjective. However, they can account for the success or failure of a program and they can be objectively measured: hence, they deserve the evaluator's attention.

Much current writing on methods of evaluation is strictly Tylerian in spirit. (See Bruner, 1966; Cronbach, 1963; and Carroll, 1965.)

Cronbach (1963) struck a distinctly traditional Tylerian chord by emphasizing the detailed analysis of course objectives, the necessity of comparing student performance with behavioral goals, and the irrelevance of comparing curricula or programs with different goals.

> The aim to compare one course with another should not dominate plans for evaluation. ....Since group comparisons give equivocal results, I believe that a formal study should be designed primarily to determine the post-course performance of a well-described group, with respect to many important objectives and side effects. (Cronbach, 1963, p. 676)

Carroll echoed Cronbach and thus, indirectly, Tyler, when he wrote:

> I would define curriculum evaluation as a process of determining whether a given curriculum attains the ends it seeks, or, rather, of determining which objectives it can attain, under what conditions, and for what kinds of pupils.... But ordinarily, curricula do not have precisely identical objectives, and it would generally be improper to compare them, because to do so would be to raise more or less philosophical questions about the comparative worth of their respective objectives. (Carroll, 1965, p )

In direct responses to these arguments against comparing educational programs, Scriven (1967, p. ) wrote: "The conclusion seems obligatory that comparative evaluation, whether mediated or fundamental, is the method of choice for evaluation problems." Two specific points which possess some similarities were advanced by Cronbach and Carroll in elaborations of their arguments: Carroll maintained that it is useless to compare curriculum A with curriculum B, because one cannot generalize from this comparison to comparisons of A with other rival curricula. Cronbach (1963, p. 676) wrote:

> At best, an experiment never does more than compare the present version of one course with the present version of another. A major effort to bring the losing contender nearer to perfection would be very likely to reverse the verdict of the experiment.

Carroll and Cronbach find fault with the comparative experimental method because it cannot achieve a purpose more properly served by research. If the comparative experiment in evaluation is open to criticism because comparing curricula A and B does not provide information about how A would compare with some unknown and unspecified curriculum C (as Carroll maintains), then this method is equally at fault for not providing any information about whether a curriculum might be produced in the future that will surpass any existing today. Furthermore, it is obvious that a comparative evaluation performed today compares only the present versions of two or more curricula.

Cronbach's statement that "a major effort" at upgrading the level of the poorer of two curricula would likely cause it to surpass its competitor in excellence is probably true. However, what effect would a similar "major effort" have on the curriculum that was initially superior? Unless some point of diminishing returns has been met in the development of the curriculum that first proved superior, major efforts on both curricula will likely leave the order of excellence unchanged on subsequent comparative evaluations.

Carroll claimed that curricula do not ordinarily have identical objectives and that to compare them raises philosophical problems about the comparative worth of different objectives. Making a choice between two competing curricula with greatly different objectives cannot help but raise philosophical questions or ethical questions or at least questions about the relative worth of certain values held by a society. Those who must make decisions affecting the adoption of curricula or innovative activities for a school are faced with resolving just such questions. I doubt that such questions can be adequately resolved and a rational decision made unless empirical data are gathered that show how well a curriculum attains its own objectives, the objectives of competing curricula, and certain cross-curricular objectives.

Many choices between competing curricula will inevitably involve philosophical questions -- questions of value. It is not the duty of the evaluator to answer these questions by himself; but he does play a vital role in cooperation with the curriculum specialist, the educational psychologist, the philosopher, and the administrator in clarifying the questions and bringing empirical data to bear on them.

One of Cronbach's major criticisms of the comparative method of evaluation was that it contributed little to understanding the curriculum:

> In an experiment where treatments differ in a dozen respects, no understanding is gained from the fact that the experiment shows a numerical advantage in favor of the new course. No one knows which of the ingredients is responsible for the advantage. (Cronbach, 1969, p. )

Scriven (1967, p. 65) answered Cronbach on this point:

> ...understanding is not our only goal in evaluation. We are also interested in questions of support, encouragement, adoption, reward, refinement, etc. And these extremely important questions can be given a useful though in some cases not a complete answer by the mere discovery of superiority.

Cronbach and Scriven are not so much at odds on this point as they are speaking at cross purposes. Scriven is clearly correct: problems of adoption of a curriculum, deciding between competing curricula, etc., require a comparative evaluation. However, Cronbach's remark seems to be addressed more to the developer of a curriculum than to the selector of one. The curriculum developer will probably find data that pinpoint the failures and successes in his materials far more valuable than data obtained from a comparison of his materials with those of a competitor. After being told that his curriculum has just won last place in a comparative experiment with his chief competitor, the typical curriculum developer would probably react in one of two ways: 1) he would maintain that the experiment was invalid, biased, and unfair; or 2) he would argue that his curriculum was not compared with its competitor on the "proper, important" objectives. In either case, he will not find such data useful in subsequent developmental work. Such data may even have the adverse effect on causing the curriculum developer to alter the objectives of his materials and begin to prize certain objectives because he can better attain them, but not for their intrinsic worth.

If the curriculum developer wants to know why and how his materials function as they do, he will not make good use of comparative data. However, comparative evaluation at some level is necessary. The critic who opposes all comparisons of curricula in favor of determining which objectives are met by which students has forgotten that there is an implicit comparison in establishing the objectives of any curriculum. No one is foolish enough to establish the objective for a curriculum in typing of "typing 10 words per minute with no more than 5 mistakes," because existing curricula are already superior to this. At some point in the evaluation of a curriculum the implicit comparisons must be revealed and subjected to test.

The comparative-noncomparative issue has been analyzed in detail because it is a point on which the Tylerian model and some other models sharply diverge. It is fair to say that the comparison of student performance with behavioral objectives and not with performance under other conditions typifies the Tyler model.

The Tyler model of evaluation has been nurtured for nearly half a century; it has almost achieved full growth. The rigidity of its defenders (see Walbesser, 1963 and 1966, for example) and the aura of orthodoxy surrounding it indicate that its potential has been realized and that in the minds of its architects it has reached full stature. We seem to have before us, then, the full-grown phenotype of the Tylerian model. What is the utility of this model? Is it appropriate to the current needs of educational evaluation?

Beginning in the second decade of the twentieth century a small proportion, approximately 4%, of the money spent on public education in the United States was collected through taxes and redistributed by the federal government. Authorized under such legislation as the Smith-Hughes and Smith-Lever Acts, these funds were expended primarily for vocational education and in the rural areas of the country. The character and rate of funds for public education channeled through the federal government changed little from 1920 to 1958. Faced with a new generation of problems and increased public concern for education, Congress reacted since then by passing the National Defense Education Act of 1958, the Elementary and Secondary Education Act of 1965, and the Education Professions Development Act of 1967. Federal support for public education nearly doubled between 1958 and 1968 (from 4% to 7% on the average for the fifty states). The thrust of these expenditures is being directed toward innovation and changing the nature of education rather than toward merely extending the services of the schools or writing new textbooks. Although the amount of federal money for innovative programs is small when measured against the total expenditure for education, it has had a profound effect on a significant number of schools. In short, the "cake of educational custom has been cracked."

Three forces are providing the motivation for the continued interest in developing models for educational evaluation. First, the proportion of funds that reach public schools through the federal government is apt to increase. It has been predicted that by 1990 approximately 50% of the cost of higher education will be met by the federal government. The ethical obligation to evaluate (to document and judge) educational programs will be heightened by this restructuring of the distribution of funds. When the entire expense of education is born by the local citizenry, feedback on the success of new programs is immediate and is acted upon quickly by those who pay the bills. However, when the cost of a new program is met with the tax dollars of anonymous taxpayers a thousand miles away, abuses and failures of the program may tend to be covered up by the community (whose attitude may be, "So what if things did not work out well, at least we got our share of the money."). Formalized evaluation requirement into Public Law 89-10 and subsequent legislation (e.g., Model Cities, EPDA) was wise.

The second and third forces that are pushing evaluation into the spotlight are the civil rights movement and teacher militancy. There is insufficient space to document the case here. However, there is evidence almost daily in the mass media that as minority groups and an aroused teaching profession lock horns with the educational establishment, each side appeals with increasing frequency to empirical evidence of the outcomes of education to resolve their difficulties. One sociologist, Dan Lortie at the University of Chicago, predicted the advent of a "certified public evaluator" who would serve a function analogous to that of the certified public accountant. His prophetic observation will be born out if action is taken on the following paragraph from the report of the National Advisory Commission on Civil Disorders (1968, p. 451):

> To increase the accountability of the public schools, the results of their performance should be made available to the public. Such information is available in some, but not all, cities. We see no reason for withholding useful and highly relevant indices of school (but not individual student) performance and recommend that all school systems adopt a policy of full public disclosure.

The public and bureaucratic clamor for something called "evaluation" caught the academics completely by surprise. Overnight, "evaluation" became a hot issue, and the central question seemed to be, "What is it?".

The academics who could first command an audience of educators when writing about evaluation were the men who had been involved in the "curriculum movement" of the 1950's. They had been thinking, speaking, and writing about evaluation as the handmaiden of curriculum research and development. Their prescription for the evaluation requirements of federal legislation was curriculum evaluation based on the Tylerian model. Models for curriculum evaluation were firmly engrained in the educational culture; they grew from the educational measurement and curriculum development moevements. They carried with them into the late 1960's the baggage of

2

objective achievement testing, taxonomies of objectives, the behavioral statement of instructional goals, etc.

It quickly became apparent that the type of evaluation called for in recent federal legislation was not just "curriculum evaluation" but something more comprehensive. What was called for was not a prescription for improving the "curriculum" (by which was generally meant printed instructional materials). A model of evaluation was needed that would determine the value (worth, benefits) of activities as diverse as a mobile learning laboratory for children of migrant workers in Washington state, a computerized system of retrieving research information for teachers in Colorado, and a legitimate theatre for underprivileged children in New Orleans.

The Tylerian model of formative curriculum evaluation is illsuited to the problems of evaluating equipment, organizational plans, staff competence, the logic of a program rationale, the goals of a program, or cost/benefit ratios. Such problems are of little interest to the Tylerian curriculum evaluator; that he should evaluate an overhead projector is inconsonant with his self-concept. However, such problems must be confronted if evaluators are to discharge their full responsibility to their clients and the patrons of education. Hence, it seems unlikely that the Tylerian model of evaluation can grow to meet the new responsibilities of educational evaluation.

### The Accrediation Model

Accrediation is the oldest type of evaluation activity. Organizations such as the North Central Association of Colleges and Secondary Schools, the American Association of Colleges for Teacher Education, and the National Council for the Accreditation of Teachers of Education seek to identify blatant deficiencies in the education of students and their teachers. When deficiencies are found, certification of programs is withheld; embargos on the graduates of censured secondary schools or colleges generally lead to a voluntary and speedy correction of substandard conditions.

The North Central Association (NCA) has a developmental history typical of that of educational accreditation agencies.* It was founded on March 29-30, 1895, by the Presidents of the Universities of Michigan, Wisconsin, Chicago, and Northwestern University and three secondary school principals. The purpose of the association was to "establish closer relations between the colleges and the secondary schools." Henceforth the association drew its members from among administrators of public and private secondary schools and colleges. NCA functioned as a debating society during the late 1890's as its membership grew to 97 institutional (58 secondary schools, 36 colleges, 3 normal schools) and 32 individual members. Between 1901 and 1910, NCA developed the accrediting policies which became its trademark. Formerly provincial college and universities were increasingly admitting a clientele that was as diverse and uneven in its secondary school preparation as it was in its geographic origins. At the NCA annual meeting in 1901, Dean Forbes of the University of Illinois spoke on "The Desirability of So Federating the North Central Colleges and Universities as to Secure Essentially Uniform or at Least Equivalent Entrance Requirements." The association responded to

---

*I have drawn heavily from Calvin C. Davis's A History of the North Central Association (1945) in this section.

Dean Forbe's address by appointing three committees on the accreditation of schools: Committee on Unit Courses of Study, Committee on High School Inspection, Committee on College Credit for High School Work.

The Committees on Unit Courses of Study and College Credit for High School Work delivered inconsequential reports at the 1902 annual meeting and were never heard of again. Thus, the association missed the opportunity to base accreditation on something akin to pupil performance and subject-matter mastery. Perhaps the timing was inopportune. The educational measurement movement was not to come about for several years; thus there was no technology of testing to draw upon.* Thereby another principle of growth is illustrated: if the raw materials do not exist in the environment, the phenotype will not reach full development regardless of the potential in the genotype.

The Committee on High School Inspection proved to be the most vigorous. Unlike the other two committees, this one could draw upon the experiences of its predecessors. State high school inspections were common in many states during the 1890's. The High School Inspection Committee proposed that secondary schools be granted membership in NCA pending satisfactory status on four standards, viz., that all teachers be graduates of NCA colleges, that teachers not teach more than four hours daily, that laboratory and library facilities of the school be adequate, and that the "general intellectual and ethical tone" of the school be adequate "as evidenced by rigid, thorough-going, sympathetic inspection." Over the years, the broad policy of the Committee on High School Inspection was interpreted and embodied in the accreditation criteria. The criteria for secondary schools in use in 1945 emphasized:

1. The "general intellectual and moral tone" of the school;
2. The school plant;
3. Instructional equipment and supplies;
4. The library and its services;
5. Financial data and personnel records;
6. Policies of the school board;
7. Organization and administration of the school;
8. Teacher qualifications (degrees, subject-matter preparation);
9. Teaching load;
10. Whether the curriculum meets pupil's needs and interests;
11. Guidance services;
12. The school as educational and recreational center for the entire community.

The accreditation criteria reflect the interests of administrators; attention is given to the processes or means of education and as opposed to its consequences on learners. The process-oriented evaluation of the early years of NCA--which shaped its future--appears to have been undertaken in a spirit of faith that altering electives, course units, teacher training requirements, and the school plant would have significant effects on the quality of learning. In developing the criteria during the first half of the century, NCA did not draw upon the burgeoning communities of behavioral scientists, psychometricians, and statisticians who ultimately played major roles in the development of other evaluation models. Opportunities for productive collaboration between these two communities on problems of evaluating learning arose at times but were never taken advantage of.

---

*Joseph M. Rice's pioneering work may have been known to some at this time, but it was probably regarded more as tendentious journalism than educational research.

As early as 1898, NCA concerned itself with the teaching of English. This must have been an interest of the more scholarly members of the association. In typical academic fashion, they responded to the question "Can uniform requirements in English be established?" with over twenty years of debate and skein of voluminous reports (while the Committee on High School Inspection passed and enforced its criteria). With the exception of secondary school accreditation--a direct outgrowth of the committee on High School Inspection--NCA was inclined to debate, issue pronouncements, and do little else.

From the beginning of NCA, instructional outcomes were wedded to the faculty theory of psychology then in vogue. It was resolved at the annual meeting in 1897 "that those studies which are best adapted to develop the faculties of the pupils should have predominant place in the several curricula...." Faculty psychology was a victim of Thorndike's associationism and Watson's behaviorism in the early 1900's. Perhaps the community of behavioral scientists and the community of accreditors found themselves so far apart on their view of the nature of the learner that collaboration was impossible.

Beginning in the early 1920's the NCA Commission on Unit Courses and Curricula sought to develop standards for evaluating the outcomes of instruction. Again this activity was undertaken largely independently of the then nascent disciplines of educational psychology and measurement. The work of this commission eventuated in a set of rather global instructional objectives: 1) impart "fruitful knowledge"; 2) develop "attitudes, interests, motives, ideals, and appreciations"; 3) develop the faculties of "memory, judgment and imagination"; 4) impart "right habits and useful skills." Apparently the activities of this Commission came to naught since the Executive Committee issued a recommendation in 1940 that accreditation activities must place more emphasis on the quality of instruction.

In seeking to determine why NCA never became substantially involved in the evaluation of the pupil performance outcomes of teaching, it would be a mistake to underrate the influence of the personalities and areas of expertise of the early members of the association. They appear to have considered themselves to be more competent to evaluate the processes rather than the products of education--just as General Motors chooses not to market lingerie.

The methodology of accreditation still borrows little from the methods of the behavioral and social sciences. Standards against which schools are measured are generally arrived at through deliberation of experts on public education. The judgment of merit of a school's program is typically arrived at directly by experts on site-visits; such judgment is not usually mediated through objective tests of student and staff performance, representative surveys of attitude and opinion, data analysis, etc. Among evaluation models, the accreditation model is incomparably superior in its use of expert judgment and comprehensive description and judgment of school administration, organization, and finance.

The Accreditation model has shown little sign of growth for several years. Like the Tylerian model, it has matured and reached the limites of its development. The Accreditation model has reached the final stage in the growth of a discipline: it has become institutionalized. When a discipline acquires an institutional identity with an administrative hierarchy, professional meetings, archival publications (e.g., the North Central Association Quarterly), etc., the probability of future revolutionary change nearly vanishes. Thus, accreditation as institutionalized in the North Central Association, the American Association of Colleges for Teacher Education, and the National Council for the Accreditation of Teachers of Education (NCATE) can be taken as the full-flowering of the Accreditation model. Will the current needs of educational evaluation be met by this model?

Evaluation methodologists can learn much from those experienced in accreditation. The comprehensiveness of school accreditations, the attention paid to the non-behavioral and non-pupil features of the school, the wisdom embodied in the elaborate checklists for observers and for faculty self-study are admirable--hopefully they can be emulated. However, although the Accreditation model "has the advantages of quick response and the utilization of the full range of the evaluator's competence, it obviously leaves much to be desired in terms of objectivity and validity, which are at best moot" (Guba and Stufflebeam, 1968, p. 11). If the Accreditation model has genetic defects-- and I believe it does--they are that its practitioners do not seek to justify empirically the standards used to judge worth and that attention to the processes of education is not balanced by attention to its consequences on learners. Minimal requirements for a school are arrived at through expert judgment that is seldom bolstered by evidence from empirical research. Schools are sometimes refused accreditation because they employ too few counselors for the number of students in the school or because their teachers do not meet certification requirements, even though one cannot point to any valid evidence that a low counselor/student ratio or uncertified teachers cause inferior education. The University of Wisconsin vs. NCATE battle of the early 1960's is an instance of an accreditation agency seeking to impose invalid and unjustifiable standards on an exemplary teacher training program.

The formulation of the Standards for School Media Programs (1969) by the American Library Association and the National Education Association is typical of the process by which accreditors have derived standards for evaluation. These standards were derived by a joint committee of 28 persons from the two above mentioned associations and with the cooperation of representatives of nearly thirty professional education associations. It is significant that none of these latter organizations has a tradition of fostering empirical research in education. The procedures by which the standards for school media programs were derived are as follows:

> After a meeting of the Advisory Board and after the first two meetings of the Joint Committee, the tentative recommendations for the quantiative standards for media

> centers in individual schools and for the unified program were presented at special sessions during the 1967 conventions of the Department of Audiovisual Instruction, the American Association of School Librarians, and the National Education Association. Reactions were invited and received. These standards were also discussed in numerous other conferences and meetings. Several thousand individuals had an opportunity to express their viewpoints during this stage of the standards. A great number indicated their opinions and suggestions. These responses were reviewed and considered carefully by the members of the Joint Committee as they compiled the text of standards.
>
> The revised draft of the standards was then submitted to over two hundred specialists in the school library and audiovisual fields (including board members of the organizations sponsoring the project, presidents of state associations, and others). Additional comments from the field were studied by the members of the Joint Committee as they continued their work on the standards in later meetings. The members of the Advisory Board then met to review the draft approved by the Joint Committee and after their recommendations had been incorporated, the standards were presented to the boards of the American Association of School Librarians and the Department of Audiovisual Instruction. (American Library Assoc., 1969, xiii, xiv.)

The Joint Committee was proud to report that a large number of persons was consulted and allowed to influence the formulation of the standards. The committee attempted to legitimize its work and throw the weight of expert consensus behind its criteria by obtaining the suggestions and implied endorsement of thousands of educators.

It is doubtful that polling educators to obtain opinions on acknowledged standards of excellence of media programs or any other enterprise can safely substitute for the empirical validation of standards. Expanding the size of the group that sets standards merely increases resources for self-deception and the protection of self-interest unless proposed standards are subjected to uncompromising attempts to demonstrate their validity with empirical data.

How well would the Standards for School Media Programs fare if put to an impartial, empirical test? Doubtless, not too well. For example, the standards hold school media programs to each of the following:

1. At lest 20 library books per student;
2. Three to six newspapers in elementary schools; six to ten newspapers in junior high and secondary schools;
3. Six tape or disc recordings per student;
4. That no more than 100 students should be seated in one area for purposes of "reading and browsing";
5. 250-400 square feet of space for storing back issues of magazines.

One may guess with little fear of contradiction that a survey employing statistical control of concommitant variables such as "wealth of the community" and "ability of students" would not show superior pupil performance on any scale for schools which save back issues of magazines over schools that don't. Indeed, such a survey would probably show only that certain schools were wasting space and money allowing back issues of magazines to pile up in the attic.

The aurhors of the school media programs standards also wished to hold schools to the requirement of employing one media expert for every 250 students in the school system and one media aide for every 2,000 students. Only the lack of a strike threat makes such recommendations different from "featherbedding" in the railroad industry. One of the most promising, innovative media programs was developed in 1969 by the Ontario Institute for Studies in Education. A large number of schools can be linked by telephone and coaxial television cables to a central media center. Within a few minutes after receiving a telephone request from a teacher, the center can telecast a film or video tape from its vast collection to an individual classroom. Such a program fails to meet most of the published Standards for School Media Programs.

However, in fairness it must be said that standards as typically derived by accreditors are not without merit. In fact, they are exemplary for their comprehensiveness and attention in detail. The danger is that standards will be unintelligently enforced. Standards are not intelligently used when they can not be demonstrated with accepted methods of proof to cause valued educational outcomes.

Evaluation will not enhance the value of an educational program if it demands conformity to standards which themselves cannot be demonstrated to lead to valued goals. The society of educational accreditors is presently estranged from the society of educational researchers who could demonstrate with empirical methods which of the accreditor's standards are valid and which are not. I see little hope for productive collaboration between these two communities. The genetic flaw in the Accreditation model will probably never be corrected; thus it will not grow into the fully useful methodology of evaluation that is needed.

## The Management-Systems Evaluation Model

Several recent attempts to organize thinking about educational evaluation have eventuated in a common class of methodologies. The models proposed by Alkin (1967, 1969) and Guba and Stufflebeam (1968; Stufflebeam, 1968) are typical of this class and will be the only models discussed here.

Guba and Stufflebeam's (1968, p. 24) definition of evaluation is quoted below in full:

> Definition: EDUCATIONAL EVALUATION IS THE (1. PROCESS) OF (2. OBTAINING) AND (3. PROVIDING) (4. USEFUL) (5. INFORMATION) FOR MAKING (6. EDUCATIONAL DECISIONS).
>
> Terms:
>
> 1. Process. A particular and continuing activity subsuming many methods end involving a number of steps or operations.
> 2. Obtaining. Making available through such processes such as collecting, organizing, analyzing, and

reporting and through such formal means as statistics and measurement.

3. *Providing*. Fitting together into systems or subsystems that best serve the needs or purposes of the evaluation.
4. *Useful*. Appropriate to predetermined criteria evolved through the interaction of the evaluator and the client.
5. *Information*. Descriptive or interpretive data about entities (tangible or intangible) and their relationships.
6. *Educational Decisions*. A choice among alternatives for action in response to educational needs or educational problems.

Guba and Stufflebeam maintain that evaluation should be viewed as the collection of information for decision-makers. Evaluation to them performs a service function of supplying data to the decision-makers charged with the conduct of the program. In their writings on evaluation, these authors focus on planning for decisions, typologies of decisions and the interrelationships of decisions in various educational contexts.

Alkin (1969, pp. 3-4) defined evaluation similarly:

> Evaluation is the process of ascertaining the decisions to be made, selecting related information, and collecting and analyzing that information in order to report summary data useful to decision makers in selecting among alternatives....
>
> The decision maker determines the questions to be asked or the decisions to be made and not the evaluator. The task of the evaluator is to determine from the decision maker the decisions for which information is required.

Alkin emphasized that evaluators should provide the decision maker with data but that they should not make judgments themselves. "The information is provided by the evaluator, but the relative weightings of the alternatives (into over-all judgments) must be made by the decision maker" (1969, p. 13). Though Alkin did not attempt any justification for this assertion, he might have sought to justify it in the manner in which the authors of *Disciplined Inquiry for Education* (1969, pp. 26-27) justified a similar statement:

> *The role of each* (decision-oriented) *study is to provide the decision-maker with information, not to tell him what to do.* ...The choice of action is the responsibility of the school executive rather than the investigator; only the executive or his advisory board is in a position to weigh the political, economic, and educational aspects of the choice.

The logic of this recommendation does not rush immediately upon one. It carries the implication, for example, that an evaluation of an educational program does not address political and economic aspects of choices. Surely any evaluation which does not do so is incomplete. It is doubtful that the subjective impressions of executives and their advisory boards can add significantly to the ability of objective data on political and economic (and sociological, etc.) questions to reduce uncertainty about the outcomes of decisions. Moreover, the position that the weightings which decision makers give sources of data are rightly and private concern of executives and their advisory boards is unacceptable. Evaluation data are worthless no matter how impeccably they are gathered if they are capriciously or unintelligently combined into value-judgments affecting decisions. The weightings which are to be applied to performance scales to determine the composite value of alternatives must be made public and must be studied explicitly by the evaluator.

A danger inheres in attempts to develop evaluation models that are models of the collection of evidence for decision making. Such models neglect two fundamental points of Scriven's definition of evaluation, viz., that the "activity consists in the...combining of performance data with a weighted set of goal scales to yield either comparative or numerical ratings, and in the justification of (a) the data-gathering instruments, (b) the weightings, and (c) the selection of goals" (Scriven, 1969, p. 40).

Decision-centered evaluation methodologist (such as Guba and Stufflebeam) argue that values are included in their thinking and their models because a decision is always the revelation of a value: if the decision-maker chooses A over B, he obviously values A more than B. They believe that values are implicit in decisions.

Guba and Stufflebeam (1968, p. 28) contend that,

> The process described as evaluation here comes much closer to the root meaning of the term, to evaluate, than does the process which currently masquerades under the name; we might argue that if a name were to be changed it ought to be that of present practice. Values come most meaningfully into play when there are choices to be made, and the making of choices is the essential act of decision-making. What we are proposing here is that the entire act of evaluation should center on the criteria to be invoked in making decisions. As we shall see, it is through the exposing of such criteria that we obtain guidance about the kinds of information that should be collected, how it should be analyzed, and how it should be reported. The term evaluation seems to be particularly suited to the process as described here, since that process makes such distinctive use of value concepts.

For a "values-centered evaluator", however, decisions would be implicit in the process of measurement against value scales, integration of measures into value-statements, and the justification of the measurement and the means of integrating the measurements; the alternative that scores highest on a weighted combination of value scales would be the preferable alternative. A decision-centered evaluation model can be applied without concentrating attention on the process by which a decision-maker integrates information into an over-all judgment.

Equating values with preferences has precedent in economics. To the economist--historically at least--the value of a product is revealed by preferences for it: if the consumer will pay $5.00 for A, then the value of A is $5.00. Such a simplistic definition of "value" treats wise and foolish evaluation equally; any $5.00 product is as valuable as any other $5.00 product. Women regularly pay $5.00 per ounce (market value) for a beauty cream, although the constituents--materials and labor--of that cream cost only 25¢ (the true value of the product). That the cream can be marketed for $5.00 is testimony to the consumers' irrational belief that expensive products must also be high-quality products. A cosmetics company once substantially lowered the price of an expensive beauty cream, which was being sold at greater than 1000% profit, only to find that sales decreased! The difference between decision-centered evaluation theorists and values-centered theorists is the difference between fixing the value of the beauty cream at $5.00 because women will pay that price for it and fixing its value at 25¢ because the total investment is a quarter. Similar illogic is common in the drug industry: some "brand name" drugs outsell identical "generic" drugs although the former cost 30 times as much as the latter. The analogy to educational evaluation is distressingly apt. Administrators have been known to choose teaching method A instead of method B, despite evaluative data to the contrary, because A is expensive. This kind of administrator's typical thoughts are "Surely all that expensive gadgetry and those priceless materials would not have been produced unless they are an improvement over old methods; the new methods must be better."

It would be satisfactory to disregard the direct assessment of value if decision-makers' preferences were always logical, rational, intelligent revelations of value. In truth, most decision-makers are perplexed by the decision-making process, and many of them rightly feel guilty and insecure about their inability to justify their decisions. Hence, it seems unwise to view evaluation as the presentation of data to decision-makers who must then make of the data what they will.

Evaluation can play many roles in an educational program; it can aid the developers by providing mastery test data, it can provide data to facilitate administration of the program, etc. However, the goal of evaluation must always be to provide an answer to all-important question: Does the program under observation have greater value than its competitors or sufficient value in itself that it should be maintained?

Guba and Stufflebeam joined earlier critics of the use of comparative experimental designs in evaluation. They concluded that "On the surface, the application of experimental design to evaluation problems seems reasonable, since traditionally both experimental research and evaluation have been used to test hypotheses about the effects of treatments. However, there are . . .distinct problems with this reasoning" (1968, p. 14). Most of the alleged problems, however, stem from Guba and Stufflebeam's unorthodox conception of the nature of comparative experimentation in the social sciences. They maintained, for example, that for comparative designs to yield valid results" . . . the treatment and control conditions must be applied and held constant throughout the period of the experiment, i.e., they must conform to the initial definitions of these conditions. The new or traditional program conditions could not be modified in process, since in that event one could not tell what was being evaluated" (1968, p. 13). Apparently they are worried about "treatments" which are so narrowly and strictly defined that they permit decision makers no freedom to adapt and "trouble shoot" while treatments are being applied. Surely such confining treatments are not requisite for valid experimental comparisons. An educational treatment may very well simply create an identifiable context within which decision makers are free to adapt the program to the exigencies of the moment. A medical researcher evaluating a drug against a placebo is free to administer other substances to control side-effects or to vary the amount of dosage in accord with his observations of the progress of remission of the disease. Such decision making does not destroy the validity of the drug-placebo comparison since it is a necessary part of the context which is being evaluated, namely the "treatment of Disease X by Drug A." Of course, the decision maker can so alter the context of the application of a treatment that the orginally defined treatment is no longer being evaluated, as for example when a medical researcher stops administering the experimental comparison does not mean that he cannot function within the context of an identifiable "treatment" without impairing the validity of a comparison.

Guba and Stufflebeam also conceived of comparative experiments as requiring that ". . . all students in the experiment must receive the same amount of the treatment to which they are assigned . . ." (1968, p. 13). Comparative experimental designs require no such thing. Here and above the authors appear to conceive a "treatment" as a fixed entity in nature like a bushel of potatoes or an M&M. A "treatment" in a comparative experiment in the social sciences is often an abstraction--a construct--with defining characteristics which create a context; the context created by the construct is all that one can evaluate. There is no need to require that the context demand that all experimental subjects receive the same amount of something. Economists ran experiments on the "negative income tax" in New Jersey in the late 1960's; persons on a negative income tax plan were compared with persons on the conventional IRS plan on variables like unemployment rate, work incentive, spending and saving habits, etc. The very essence of the negative income tax is that its amount varies from person to person, yet I know of no one who claims that the comparison is thereby invalidated. Indeed, not all subjects need even receive the same thing, as when we evaluate individualized instruction.

Guba and Stufflebeam (1968, p. 14, 15) claimed that the application of comparative experimental design to evaluation problems ". . . conflicts with the principle that evaluation should facilitate the continual improvement of a program" and that ". . . it is useful for making decisions after a project has run full cycle but almost useless as a device for making decisions during the planning and implementation of a project." It is reassuring that the utility of comparative experimental design for "end-of-project" decisions is acknowledged by two more authors. The critical points which Guba and Stufflebeam raised were resolved earlier by Scriven's distinction between formative and summative evaluation when Cronbach (1963) raised the same points.

Guba and Stufflebeam faulted comparative experimental design because of the "near impossibility" of controlling or eliminating "confounding variables" through randomization or otherwise. Cronbach raised the same point earlier: "Any failure to equate the classes taking the competing courses will jeopardize the interpretation of an experiment and such failures are almost inevitable" (1963, p. ). One does not seek to "equate" comparison groups; such equation of groups is an impossibility recognized early in the history of experimental design. In comparative experimental design, groups are made "randomly equivalent" -- which isn't equivalent at all--and post-experimental differences are inspected to reveal whether they are small enough to be attributed to the original random assignment or whether a treatment effect must be postulated to account for the large difference. Thus, valid experimental comparisons are not impossible just because experiments cannot perfectly equate groups. Valid, probabilistic comparisons are possible, as the growing number of well-designed comparative experiments in education demonstrates. It is true that valid experimental designs are difficult and expensive to implement; but educational researchers and evaluators have yet to learn that such designs are usually worth the cost.

Finally, Guba and Stufflebeam (1968, p. 16) wrote that "A fourth problem inherent in the application of conventional experimental design is the possibility that while internal validity may be gained through the control of extraneous variables, such an achievement is accomplished at the expense of external validity." This assertion possesses the symmetry which sounds the ring of truth to the reader who is untutored in the methods of empirical research. Internal and external validity are not bipolar opposites. Designing experiments which evidence both types of validity to a high degree is simply a set of technological problems in instrumentation, data collection, and statistical analysis (see Bracht and Glass, 1968).

The Tylerian and the Management-Systems models stress certain roles of evaluation rather than striving to attain its goal. Traditional "curriculum evaluation" models have emphasized playing various roles in the development or operation of a program; in some instances the proponents of these methods have even argued against attempting to achieve the goal of evaluation. The Management-Systems evaluators also appear to be more concerned with playing a role supportive of administrators than with adjudicating questions of value. Being of assistance to the program personnel -- so they may better conduct their business -- is a proximate aim of evaluation; the ultimate aim of an evaluation is to decide questions of worth. An evaluator's rendering of judgment on the composite value of an educational program poses a threat to teachers and administrators, whom he might live with more amicably "in a service capacity." Nevertheless, he is obliged to make the judgment; he cannot safely shirk the obligation.

## The Composite-Goal Model

The model of evaluation which I have chosen to call the "Composite-Goal model" is due to Scriven (1967).

Scriven (1967, p. 40) defined evaluation as follows:

> Evaluation is itself a methodological activity which is essentially similar whether we are trying to evaluate coffee machines or teaching machines, plans for a house or plans for a curriculum. The activity consists simply in gathering and

> combining of performance data with a weighted set of goal scales to yield either comparative or numerical ratings, and in the justification of (a) the data-gathering instruments, (b) the weightings, and (c) the selection of goals.

Scriven's definition of evaluation (in which composite criteria of worth are emphasized) yields the unique model of evaluation referred to here as the Composite-Goal model. In my opinion the Composite-Goal model of evaluation is the only one discussed here with the potential to grow into a fully useful methodology of evaluation.

The potential utility of the Composite-Goal model derives from its focus on the direct assessment of worth (which distinguishes it from the Management-Systems model), its concern with the justification of the valued criteria and goals (which distinguishes it from the Accreditation model), and its comprehensive character that will permit its application in the diverse contexts now calling for educational evaluation (which distinguishes it from the Tylerian model). The Composite-Goal model is the only model discussed here which genuinely "models" the act of evaluation. The process by which one rationally arrives at a defensible assessment of the worth of an enterprise or an object is well typified by Scriven's tripartite definition of evaluation. The Accreditation model is simply inadequate for producing comprehensive and defensible value judgments. The Tylerian and Management-Systems models are fine models, to be sure. However they do not "model" the process of evaluation; rather they are models of curriculum development and program administration, respectively. The greatest growth of the Composite-Goal model is still to be realized. If the model is to achieve full growth and utility, there are several features contained in the definition of the Composite-Goal model for which practical technologies must be developed.

## The Nuture of the Composite-Goal Model of Evaluation

The discussion on how the growth of the Composite-Goal model may be fostered can profitably center around Scriven's definition of evaluation. Evaluation methodologists have not yet devised many of the techniques necessary to implement the Composite-Goal model. Indeed much remains to be learned about every element of Scriven's definition: a) what data and at what level of generality-specificity should performance data be gathered? b) how should data be weighted into summative composites to yield ratings of worth? c) how can one justify the data-gathering techniques, the weights in composites and the selection of goals? Each of these questions calls for evaluation techniques still not discovered. My purpose in this section is to refine the questions and indicate something about how the needed techniques might be found.

### A. Gathering Data

Two unsolved problems of evaluative data collection involve determining the appropriate level of specificity at which the most meaningful data lie and establishing priorities for the collection of data judged to be meaningful.

Generality-Specificity of Data. Anything as complex as an educational program can be examined at innumerable levels of specificity (Krathwohl, 1965). Evaluators are advised to heed a vast assortment of data. They are warned that anything that feeds into a program (antecedents), happens during it (transactions), and results from it (outcomes) may prove to be critical to the success of the program. They are also told that it is vital to consider not only what happened (observations) but what should have happened (intents). Evaluators are not told, however, the level of generality-specificity at which it is wisest to state intentions and make observations. Lacking such guidelines, evaluators may fail to record the essential character of the programs they evaluate. Tyler (1966) identified the problem of determining the appropriate level of specificity for the statement of instructional objectives as the most vexing that instructional researchers now face. He found that the behavioral objectives are sometimes stated so specifically that generalizations of specific facts are never consciously taught and, hence, are not learned. Observations of an educational program can be molecular when the telling data are at a higher level of generality.

The following fantasy is an illustration of how observation must be guided by a methodology if it is to avoid irrelevance. A Martian was sent to Earth to observe its inhabitants. Upon his return to Mars he filed this report with his superiors: "The planet Earth is inhabited by billions upon billions of winged and six-and eight-legged creatures living in close and fascinating interaction. Their short lives are free from external dangers except for infrequent intrusions upon their dominion by a huge, fleshy creature of which there are no more than 3.5 billion on the entire planet." The Martian did indeed make some perceptive observations, but we--in our egocentric way--think that he missed the point of the planet because he looked at the wrong things.

At what level should the evaluator look for the significant phenomena in an educational program? Should "intended transactions" be a minute-by-minute lession plan or a weekly calendar of general topics and activities? Should he measure the cognitive outcome "knowledge of the animal phyla" or the outcome "identification of the species, genus, and phylum of the Tasmanian Devil"? (Attempts to dodge such questions by claiming that they must be answered by the program personnel and not by the evaluator are contrary to an honest and productive conception of the evaluation activity.)

Evaluation methodologists have yet to suggest any means for determining whether one should observe general or specific phenomena. Without the guidance of explicit methodology, too many efforts to evaluate will become either absurdly reductionistic or worthlessly global.

Priorities on Evaluation Data. Evaluation methodologists have adopted the notion (explicitly and by example) that practically all data merit collection and analysis. What is surprising and impressive in recent writings on evaluation methodology is the number and variety of variables and events that are considered worthy of observation. According to Stake (1967) the data of the evaluation effort are descriptions and judgments of the antecedents, transactions, and outcomes and the contingencies among them. Stake considers an extraordinarily wide range of phenomena as elements of the evaluation "data matrix".

Recent writings on evaluation have stimulated a salubrious broadening of vision and increased alertness to the existence of a great quantity of potentially valuable data that were formerly overlooked or considered to be peripheral. In an important sense, the "opening up" of the data matrices of evaluation was partly a reaction against the narrow and ruthlessly enforced priorities placed on data by "hard-headed behaviorists". This truculent breed of behaviorist regards performance data on behavioral objectives as the only type relevant to the evaluation of instruction. Evaluation methodologists have been hesitant to place priorities on evaluation data, perhaps because they fear that the near-sighted attacks on evaluation problems which characterized recent decades would quickly be reinstated under a new system of priorities. There is no need to fear a new generation of narrow and unnecessarily limited evaluation attempts, however, if a methodology for generating priorities for data is forthcoming rather than a new system of priorities.

A "decision" embodies two or more alternative actions: "Making the decision" is simply choosing one alternative. Considerations of impending decisions will determine in large part the data to be gathered and how they will be analyzed. Corresponding to each decision that must be made are data relevant to it. Establishing priorities on the decisions that must be made (e.g., ranking the decision from "most" to "least" in need of empirical data") is equivalent to establishing priorities on data to be gathered. Priorities might be established on the basis of the need to bring empirical data to bear on a decision. A system of priorities on the collection of evaluation data could be determined by the impending, anticipated decisions that will be faced and by the necessity to make some provisions for unanticipated decisions that are certain to arise in the course of events.

A temporarily workable methodology for establishing priorities on the collection of evaluation data might involve 1) the costs of gathering different data, 2) estimates of the prior probabilities that each alternative embodied in a decision will be supported by the data--if they were to be gathered, and 3) the costs of implementing each of the alternatives of the decision.

The nature of the three components of this embryonic methodology is clarified below; I have given illustrations of how each would act independently in determining priorities on data collection.

1. Costs of gathering different data.

   Suppose that all factors other than the differential costs of gathering evaluation data are equal. In this instance, evaluation resources are best spent by making the maximum number of decisions (because the various decisions are assumed to be equally costly, equally valuable, and our prior expectations are that the data gathered on any decision are equally likely to support each alternative of the decision).

2. Prior probabilities that each alternative embodied in a decision will be supported by the data if they are gathered.

   Imagine that all things are equal except the following: Decision I has two alternatives: A and B. The prior

probability--perhaps it is the evaluator's personal probability--that the data if gathered will support A is Prob (A) = .90; Prob (B) = .10.
Decision II has two alternatives: C and D. The prior probabilities that the relevant data would support C is judged to be Prob (C) = .50. Hence, Prob (D) is .50.

Therefore, the outcomes of gathering data can be fairly confidently guessed for Decision I, but not for Decision II. Obviously, then, the priority on gathering data bearing on Decision II is higher than the priority on gathering data to make Decision I. If our estimates of the prior probabilities have high validity, Decision I can be made without gathering empirical data.

3. The costs of implementing each alternative of the decision.

Any decision comprises two or more alternatives for which costs of implementation can be estimated. Alternatives A and B of Decision I might cost $10,000 and $11,000 respectively, if implemented. Alternatives C and D of Decision II might cost $1,000 and $5,000 respectively if implemented. If only one decision can be made with evidence and the other must be decided by a coin flip, which decision should be based on empirical data? The answer depends on not only the costs of the alternatives, but also on the benefits that would result from implementing each alternative and the "loss" of implementing inferior alternatives.

In spite of the apparent promise of such rudimentary decision methodologies and the ease with which they can be formulated, all of them probably assume that too much is known _a priori_ to be of much immediate application to educational evaluation. Assuming that all alternatives of a decision are known before any data are gathered is even too simplistic for the present state of educational technology. Nonetheless, intrepid investigators armed only with crude heuristics may win battles while the meek wait for faultless techniques. Boulding (1969, pp. 7-8) urged the use of current first-approximations to fully mature cost-benefit schemes:

> The whole idea of cost-benefit analysis, for instance, in terms of monetary units, say 'real' dollars of constant purchasing power, is of enormous importance in the evaluation of social choices and even of social institutions. We can grant, of course, that the 'real' dollar, which is oddly enough a strictly imaginary one, is a dangerously imperfect measure of the quality of human life and human values. Nevertheless, it is a useful first approximation, and in these matters of evaluation of difficult choices it is extremely useful to have some first approximation that we can modify. Without this, indeed, all evaluation is random selection by wild hunches.

Despite widespread curiousity about cost-benefit analysis, cost-utility analysis, program planning and budgeting, etc., such methods have influenced education only at a macro-economic level. Evaluation methodologists have been little concerned with the assessment of costs and the relationships of costs to utilities. The problem of establishing priorities on the collection of evaluation data could initiate a greater concern with costs and resource allocations.

### B. Weighting Data in Composites

Almost all summative evaluation is comparative. Summative evaluation usually involves the measurement of competing programs on performance or goal scales and the integration of the data into a conclusion of superiority for one program. Evaluation methodologists have given practically no attention to the methods of integrating information into a summative judgment. Scriven wrote that the process of combining measures of valued performance is a process of summing weighted goal or performance scales; the program receiving the highest total score would presumably be preferred. The weights for the summated rating probably derived from human judgment and statistical properties of the scales. Evaluation methodologists can draw upon a highly developed psychometric theory of the measurement of judgment and the integration of information into weighted composites. In the weighted-sum model, a compensatory view of relative performance is taken. If program A is inferior to B on scale 1, it may still be preferred to B over-all because A's meritorious performance with respect to scale 2 compensates for its inferiority on 1. However, the "weighted-sum model" is just one of several conceivable models of the integration of data into summative conclusions. There are non-compensatory models in which deficiencies on one scale cannot be redeemed by extraordinary performance on other scales. With such non-compensatory models, the integration of data into a summative decision might be a simple matter of choosing the program that is superior--regardless of the degree of superiority--on the greatest number of unweighted scales.

Many decision-makers adopt a decision model based on a mini-max principle. The mini-max principle embodies the wisdom of caution: it is better to avoid disaster than to make even large gains. Rather than maximizing his accomplishments, the mini-max decision-maker wishes to minimize the chances of suffering a maximum loss. Although curriculum A is greatly superior to B on almost all scales, the mini-max decision-maker may choose B because his teachers' dissatisfaction with the amount of preparation time required for A bodes a mutinous uprising which he feels must be avoided at all costs.

Management science has recently adopted Bayesian decision models for applications in business. These models are a meld of information and human judgment into decision-making strategies.* Evaluation methodologists might be able to advance their discipline significantly through the study of models for the integration of information and judgment into summative decisions.

If the methods for combining information into summative value-statements are not understood, the process will be governed by prejudice, caprice, or irrationality. Understanding can be the beginning of control and improvement of this vital activity.

---

*The principal reference in this new field is Schlaifer (1959).

## C. Justification of Data-gathering Instruments, Weights in Composites, and the Selection of Goals.

### 1. Justification of Data-gathering Instruments.

Decades of measurement research in education, psychology and sociology have produced well-articulated theories of measurement and a variety of useful data-gathering instruments. Psychometric theories of reliability and criterion and construct validity contribute greatly to the practice of evaluation. Yet there are still unresolved problems in the utilization and justification of human judgment as data in evaluation. Scriven (1967) and Stake (1967) legitimized the use of human judgment in evaluation. Increasingly, evaluators are recognizing that--contrary to scientific canons of "objectivity"--people can be the most efficient and effective information processors. Evaluation has profited most in this decade from evaluators' new-found willingness to exploit the incomparable ability of humans to collect, store, and integrate information and render judgments.

Unfortunately, evaluation methodologists have done little beyond arguing that judgments are valuable data and that psychometric theory can help describe them. Psychometrics, however, contributes little more to the study of the judgmental process than methods for measuring judge-agreement and describing judgmental points of view. Evaluators presently have no methodology for assessing the pre-eminent quality of judgments, namely, their validity.

Perhaps the validity of judgment can be most enhanced by seeking those few individuals whose perspicacity of circumstances uniquely qualify them to render valid judgments. An intelligent executive shares the evaluators' need for perceptive judgments. He is relatively unconcerned with measuring judge "homogeneity". Indeed, he anticipates discordant judgments. The executive's job is not to resolve differences of opinion or to make judgments homogeneous, but rather to discern whose judgment is good or bad on a particular question. In the simplest social organizations, the participants quickly determine the validity of the information that any other person can supply. From the family to the corporation, constituents interact to determine who knows what about what. In most instances, a young child in a family is regarded as a dubious judge of the best color to paint the living room or of the possibility that the basement is inhabited by spooks, but an excellent judge of the state of his own hunger or dryness. An executive's job is to determine who can supply the best knowledge to serve as a basis for his decision. (One of the executive's greatest problems is that as he rises in the organizational hierarchy, he loses touch with [interacts less with] the technicians whose information he must use. Without intimate knowledge [constant testing, etc.] of his employees, he soon loses a feel for which technician is to be believed on a particular problem.) An analogy with evaluation points up the problem of determining whose judgment is worth heeding and whose is not. This is a far more difficult question to answer than whether judges A and B hold similar opinions. However, those who fail to address the question of the validity of judgment rob the judgmental process in evaluation of its power and importance.

There are, however, important instances in which the validity of judgmental data (i.e., their truth or fidelity) is irrelevant. Judgments may be studied solely as concomitants or as predictors of future actions. If his expressed judgments predict some important future behavior, it is quite useless to criticize gathering a potential decision-maker's judgments because they are "subjective," "mere impressions," etc. For example, if a principal's pro or con feelings about the disruptive character of a new curriculum predict with 90% accuracy the adoption or rejection of that curriculum, one may never care whether the principals are truly competent judges of "disruptive curricula." Their abilities as judges of such phenomena aside, an important and functional concomitance has been observed.

The agreement of a group of judges need not always concern evaluators; nor is the validity of judgment always of concern. The reliability of judgmental data may be considered apart from their validity. However, evaluators presently have only a little methodology borrowed from psychometrics to apply in the study of judge agreement and practically no methodology for studying judge validity.

## 2. Justification of Weights in Composites

At the heart of the Composite-Goal evaluation model lies the problem of combining data on different performance scales into a single rating of merit. Regardless of the methods of combining performance data that may be chosen, an evaluator will eventually face an elemental problem of equating performance on differing criteria. For example, when defining a composite measure of value of a secondary-school mathematics curriculum, should mastry of problem-solving skills be weighted twice or half as much as memory of facts? That evaluation methodologists seldom take such legitimate questions seriously testifies to the lack of a technology to deal with such important problems.

As the technology of curriculum development improves, the problem of how to assign relative weights to criteria in forming a composite value scale will become increasingly important. An improved curriculum development technology should allow curriculum writers to achieve the objectives they wish. The typical empirical evaluation of the future might simply confirm that each curriculum achieved its objectives--some of which were unique and some common to all curricula compared. The true determination of value will then become the weighting of performance data into a composite scale.

Perhaps the answer to the weighting problem lies in the discovery of a fundamental unit of utility (benefit or value) with cross-objective validity. The present need for a unit to measure educational value is reminiscent of the growth of descriptive linguistics.

Linguistics progressed very little for years because the variety of verbal utterances defied codification. Linguistic studies were revolutionized by the definition of the "phoneme" as the smallest unit that discriminated at least two spoken words. Thereafter, linguistics flourished. Research on the psychology of sleep was revivified by the discovery of rapid eye movement, REM. We may be approaching a similar stage in the development of evaluation in which the discovery of a cross-curricular unit of utility will permit the genuine assessment of value of educational programs and pump new life into stalled methdologies of evaluation.

### 3. Justification of Goal Selection

Unlike the Tylerian model in which goals are accepted without question or the Accrediation model in which goals are judged but sometimes invalidly, the Composite-Goal model attends particularly to whether those goals of a program ought to be sought. Scriven (1967, p. 52) wrote that ". . . evaluation proper must include, as an equal partner with the measuring of performance against goals, procedures for the evaluation of the goals." However, Tyler (1951, p. 48) did not lay emphasis on the evaluation of goals themselves: " 'Evaluation' designates a process of appraisal which involves the acceptance of specific values and the use of a variety of instruments of observation, including measurement, as bases for value-judgments."

Suppose that the developer of a ninth-grade social studies curriculum in Iowa decides to shorten by half a year-long unit on "Modern World Problems" and to substitute a unit on Iowa history. The Tylerian evaluator would be expected to assist the curriculum developer by refining the statement of the objectives of the new unit and by providing the developer with evidence of the success of his materials. The accreditor might register an objection to incorporating the unit into the curriculum because it might lead to an Iowa history requirement for certification of teachers. The Systems-Management analyst might attempt to determine the data that the curriculum developer would need to institute his materials in the schools. However, one might expect the evaluator who uses the Composite-Goal model to determine whether Iowa ninth-graders ought to study a semester unit on Iowa history; he may discover that 85% of ninth-grade students who would be involved leave the state by the age of 22 and never return. He may be led to the conclusion that in such a mobile society, the provincialism of a full semester devoted to Iowa history cannot be justified. In stressing the necessity that evaluation address the question of goal justification, Scriven wrote:

> Of course, if we do not know that (and usually how) . . . performance bears on merit it is a travesty to refer to the measurement of it as evaluation: and exactly this travesty is involved in a great deal of curriculum evaluation where no defensible conclusions about merit can be drawn from the kind of data that is so earnestly gathered. Good conceptual analysis (of the relevant concept of merit

> in terms of the qualities involved in it) and good experimental design are essential presuppositions of any performance-testing in the evaluation process. (Scriven, 1966, pp. 6, 7)

It is surprising how many scholars still maintain that science has no business addressing questions of value. When a noted psychometrician takes pen in hand, the reader is treated to a modern statement of de gustibus non disputandum gratuitously generalized to scientific research and its applications:

> Science, it is often pointed out in discussions of its methodology and objectives, is concerned only in discovering functional relationships among variables, without being concerned whether the variables themselves or the functional relationships are worthwhile. It cannot concern itself with moral, ethical, or social values, except as it may attempt to define variables in these areas and discover relationships among them . . . This does not mean that scientists as persons need not or should not be concerned with value judgments and with moral and ethical considerations. It only means that these considerations are not appropriate preoccupations of scientific methodology or procedure. It is unfortunate that this distinction has not been made more empirically and explicitly. Many persons find it difficult to disentangle value concepts from scientific concepts. If value judgments are made, and objectives or goals are set in terms of these value judgments, then it is the legitimate role of science to develop, formulate, or investigate methods for achieving these goals, but science cannot tell whether the goals should be achieved. Scientific methods may determine whether the attainment of certain objectives will facilitate the realization of other objectives, but they cannot in their very nature say whether the objectives are good or bad, except insofar as they promote the attainment of other objectives. (Horst, 1966, p. 335)

Few philosophers of science would agree with Horst. The modern position on the relationship of science to values is reflected in the statement of the task Kaplan set for himself in the tenth chapter of The Conduct of Inquiry (1964, p. 373):

> The Thesis I want to defend is that not all value concerns are unscientific, that indeed some of them are called for by the scientific enterprise itself, and that those which run counter to scientific ideals can be brought under control--even by the sciences most deeply implicated in the value process.

The reader who remains skeptical is referred to Glanville Williams' The Sanctity of Life and the Criminal Law, a masterful, logical and scientific, empirical analysis of the moral and social aspects of birth control, sterilization, artificial insemination, abortion, suicide, and euthanasia. If philosophers and social scientists can approach closure on questions as profound as these, then educationists need not become disheartened over the difficulty of assessing the relative values to society of a few curricula.

Educational writing is shot through with moralizing about this or the other curriculum or method of instruction. Determination of the relative merits of "expository" and "discovery" teaching (the two positions represented for example by David Hawkins and David Ausubel, respectively) must rest on an analysis of the definitions of the two terms and on empirical, longitudinal studies of the effects of each method to be on retention of knowledge, interest, motivation, career plans, personality, etc. Current debates about the superiority of discovery teaching over expository teaching flounder for want of serious attempts to analyze the terms logically and to gather the telling empirical data.

The justification of educational goals will undoubtedly draw upon both logical and empirical analyses. Philosophers can contribute greatly to the solutions of the problem of justifying goal selection by studying the logical consistency of program goals with program philosophies or rationales and the larger philosophies that guide education. Academicians can be asked whether goals relevant to their discipline can be justified. For example, a biologist is most competent to judge whether Lysenkoism should be taught for its value as a scientific inquiry in a high-school biology course. Social scientists can, perhaps, contribute most among all scholars to the solutions of the problems of goal selection.

Psychology will often be highly relevant to the justification of a curriculum goal. Consider, as an example, the American Association for the Advancement of Science (AAAS) elementary science curriculum. The writers of this curriculum viewed science as a collection of a small number of highly transferable "processess"--the "scientific methods" in a real sense. The AAAS materials seek to impart these inquiry skills to the pupil; the context of their application, i.e., the "content" of the science curriculum, is held to be largely unimportant. Some critics have attacked the AAAS curriculum; they claim that is rests on an obsolete, 19th century faculty psychology. They argue that psychological research has shown that the mind cannot correctly be regarded as a collection of faculties or abilities that can be strengthened through use and then applied in a variety of settings. Whether the AAAS materials are based on such conception of the learner and whether such a conception is without merit as a theory of behavior are questions that psychologists are uniquely qualified to answer. The answer would certainly bear on the justification of the "process approach" taken by the AAAS curriculum materials.

The need is great for educational research that would justify the selection of educational goals. We lack even the most rudimentary data--of a longitudinal sort--on the relative retention of knowledge and interests. How are we to know, then, whether a curriculum developer chooses wisely when he concerns himself with engendering interest in mathematics instead of teaching more mathematical content? If longitudinal surveys show that mathematical content is forgotten within five years after terminating formal schooling, but that an interest in mathematics perseveres and leads to further informal study and accepting attitudes of science, then the curriculum developer's selection of goals is probably justified.

It is apparent that educational evaluation will be greatly dependent upon science and other areas of scholarship for knowledge to settle questions of the justification of goal selection.

## Conclusion

Like any complex human fabrication, evaluation methodology has no real genotype; its only genotype is a plan for its future growth in the minds of its builders.

## REFERENCES

Alkin, Marvin C. Towards an Evaluation Model: A Systems Approach. Working Paper No. 4. Los Angeles: Center for the Study of Evaluation of Instructional Programs (CSEIP), University of California, Aug., 1967.

Alkin, Marvin C. The Use of Behavioral Objectives in Evaluation: Relevant or Irrelevant? Paper presanted to the Eighteenth Annual ETS Western Regional Conference on Testing Problems; San Francisco, Calif., May 9, 1969.

American Library Association and National Education Association. Standards for School Media Programs. Chicago: American Library Association (50 E. Huron, St., Chicago, Ill. 60611), 1969.

Boulding, Kenneth E. Beyond Economics. Ann Arbor: University of Michigan Press, 1968.

Boulding, Kenneth E. Eccnomics as a moral science, American Economic Review, 1969, 59, 1-12.

Boulding, Kenneth E. The Organizational Revolution. New York: Harper & Brothers, 1953.

Bracht, Glenn H. and Glass, Gene V. The external validity of experiments. American Educational Research Journal, 1968, No. 4, 437-474.

Bruner, Jerome S. Toward a Theory of Instruction. Cambridge: Harvard University Press, 1966.

Carroll, John B. "School learning over the long haul." Chapter 10 in Krumboltz, John D. (ed.) Learning and the Educational Process. Chicago: Rand McNally, 1965.

Cronbach, Lee J. Course improvement through evaluation. Teachers College Record, 1963, 64, 672-683.

Cronbach, Lee J. and Suppes, Patrick. Research for Tomorrow's Schools: Disciplined Inquiry for Education. New York: Macmillan, 1969.

Davis, Calvin O. A History of the North Central Association. Ann Arbor, Michigan: North Central Association of Colleges and Secondary Schools, 1945.

Guba, Egon G. and Stufflebeam, Daniel L. Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action. An address delivered at the Second National Symposium for Professors of Educational Research: Boulder, Colorado; November 21, 1968.

Horst, Paul. Psychological Measurement and Prediction. Belmont, California: Wadsworth Publishing Company, 1966.

Kaplan, Abraham. The Conduct of Inquiry. San Francisco: Chandler, 1964.

Krathwohl, David R. Stating objectives appropriately for program, for curriculum, and for instructional materials development. Journal of Teacher Education, 1965, 12, 83-92.

Report of the National Advisory Commission on Civil Disorders. New York: Bantam Books, 1968.

Schlaifer, Robert. Probability and Statistics for Business Decisions. New York: McGraw-Hill, 1959.

Scriven, Michael. "The methodology of evaluation" in Stake, Robert E. (ed.), AERA Monograph Series on Evaluation, No. 1. Chicago: Rand McNally, 1967, 38-89.

Scriven, Michael. Value claims in the social sciences. Boulder, Colorado: Social Science Education Consortium, Publication No. 123, 1966.

Smith, Eugene R. and Tyler, Ralph W. Appraising and Recording Student Progress. New York: Harper and Row, 1942.

Stake, Robert E. (ed.) AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Stake, Robert E. The countenance of educational evaluation. Teachers College Record, 1967, 68, 523-540.

Stake, Robert E. Generalizability of program evaluation: the need for limits. Educational Product Report, 2, No. 5, February 1969, 39-41

Stufflebeam, Daniel L. A depth study of the evaluation requirement, Theory into Practice, 1966, 5, 121-133.

Stufflebeam, Daniel L. Evaluation as Englightenment for Decision-Making. Columbus: Evaluation Center (1712 Neil Avenue), Ohio State University, January 19, 1968.

Tyler, Ralph W. The functions of measurement in improving instruction. In Lindquist, E. F. (ed.). Educational Measurement, Washington, D.C.: American Council on Education, 1951.

Tyler, Ralph W. New dimensions in curriculum development. Phi Delta Kappan, 1966, 48, 25-28.

Walbesser, Henry H. Curriculum evaluation by means of behavioral objectives. Journal of Research in Science Teaching, 1963, 1, 296-301.

Walbesser, Henry H. Science curriculum evaluation: observations on a position. The Science Teacher, 1966, 33.

Williams, Glanville. The Sanctity of Life and the Criminal Law. New York: Alfred A. Knopf, 1968.

[Reprinted from TEACHERS COLLEGE RECORD, Vol. 68, No. 7, April, 1967]

TEACHERS COLLEGE RECORD
VOLUME 68 · NUMBER 7
APRIL 1967

ROBERT E. STAKE
UNIVERSITY OF ILLINOIS

# The countenance of educational evaluation

President Johnson, President Conant, Mrs. Hull (Sara's teacher) and Mr. Tykociner (the man next door) are quite alike in the faith they have in education. But they have quite different ideas of what education is. The value they put on education does not reveal their way of evaluating education. Educators differ among themselves as to both the essence and worth of an educational program. The wide range of evaluation purposes and methods allows each to keep his own perspective. Few see their own programs "in the round," partly because of a parochial approach to evaluation. To understand better his own teaching and to contribute more to the science of teaching, each educator should examine the full countenance of evaluation.

Educational evaluation has its formal and informal sides. Informal evaluation is recognized by its dependence on casual observation, implicit goals, intuitive norms, and subjective judgment. Perhaps because these are also characteristic of day-to-day, personal styles of living, informal evaluation results in perspectives which are seldom questioned. Careful study reveals informal evaluation of education to be of variable quality—sometimes penetrating and insightful, sometimes superficial and distorted.

Formal evaluation of education is recognized by its dependence on checklists, structured visitation by peers, controlled comparisons, and standardized testing of students. Some of these techniques have long histories of successful use. Unfortunately, when planning an evaluation, few educators consider

---

**Dr. Stake, who is Associate Director of CIRCE (Center for Instructional Research and Curriculum Evaluation) at Illinois, here takes an innovative and suggestive approach to the problem of formal evaluation. He offers a conceptual background for developing a plan of evaluation of educational programs rather than educational products; and, in doing so, he makes the significant point that "the two basic acts of evaluation" are description and judgment, both of which are essential if educational programs are to be understood. Drawing attention to the need for data banks documenting information on antecedent conditions, transactions, and intents—as well as "goals" and "objectives"—he makes, we believe, an invaluable contribution to the clarification of guidelines and the rational choice of programs for public schools.**

Reproduced by permission.

even these four. The more common notion is to evaluate informally: to ask the opinion of the instructor, to ponder the logic of the program, or to consider the reputation of the advocates. Seldom do we find a search for relevant research reports or for behavioral data pertinent to the ultimate curricular decisions.

Dissatisfaction with the formal approach is not without cause. Few highly-relevant, readable research studies can be found. The professional journals are not disposed to publish evaluation studies. Behavioral data are costly, and often do not provide the answers. Too many accreditation-type visitation teams lack special training or even experience in evaluation. Many checklists are ambiguous; some focus too much attention on the physical attributes of a school. Psychometric tests have been developed primarily to differentiate among students at the same point in training rather than to assess the effect of instruction on acquisition of skill and understanding. Today's educator may rely little on formal evaluation because its answers have seldom been answers to questions *he* is asking.

## Potential Contributions of Formal Evaluation

The educator's disdain of formal evaluation is due also to his sensitivity to criticism—and his *is* a critical clientele. It is not uncommon for him to draw before him such curtains as "national norm comparisons," "innovation phase," and "academic freedom" to avoid exposure through evaluation. The "politics" of evaluation is an interesting issue in itself, but it is not the issue here. The issue here is the *potential* contribution to education of formal evaluation. Today, educators fail to perceive what formal evaluation could do for them. They should be imploring measurement specialists to develop a methodology that reflects the fullness, the complexity, and the importance of their programs. They are not.

What one finds when he examines formal evaluation activities in education today is too little effort to spell out antecedent conditions and classroom transactions (a few of which visitation teams do record) and too little effort to couple them with the various outcomes (a few of which are portrayed by conventional test scores). Little attempt has been made to measure the match between what an educator indends to do and what he does do. The traditional concern of educational-measurement specialists for reliability of individual-student scores and predictive validity (thoroughly and competently stated in the American Council on Education's 1950 edition of *Educational Measurement*)[1] is a questionable resource. For evaluation of curricula, attention to individual differences among students should give way to attention to the contingencies among background conditions, classroom activities, and scholastic outcomes.

This paper is not about what should be measured or how to measure. It is background for developing an evaluation plan. What and how are decided later. My orientation here is around educational programs rather than educational products. I presume that the value of a product depends on its program of use. The evaluation of a program includes the evaluation of its materials.

The countenance of educational evaluation appears to be changing. On the pages that follow, I will indicate what the countenance can, and perhaps, should be. My attempt here is to introduce a conceptualization of evaluation oriented to the complex and dynamic nature of education, one which gives proper attention to the diverse purposes and judgments of the practitioner.

Much recent concern about curriculum evaluation is attributable to contemporary large-scale curriculum-innovation activities, but the statements in this paper pertain to traditional and new curricula alike. They pertain, for example, to Title I and Title III projects funded under the Elementary and Secondary Act of 1966. Statements here are relevant to any curriculum, whether oriented to subject-matter content or to student process, and without regard to whether curriculum is general-purpose, remedial, accelerated, compensatory, or special in any other way.

The purposes and procedures of educational evaluation will vary from instance to instance. What is quite appropriate for one school may be less appropriate for another. Standardized achievement tests here but not there. A great concern for expense there but not over there. How do evaluation purposes and procedures vary? What are the basic characteristics of evaluation activities? They are identified in these pages as the evaluation acts, the data sources, the congruence and contingencies, the standards, and the uses of evaluation. The first distinction to be made will be between description and judgment in evaluation.

The countenance of evaluation beheld by the educator is not the same one beheld by the specialist in evaluation. The specialist sees himself as a "describer," one who describes aptitudes and environments and accomplishments. The teacher and school administrator, on the other hand, expect an evaluator to grade something or someone as to merit. Moreover, they expect that he will judge things against external standards, on criteria perhaps little related to the local school's resources and goals.

Neither sees evaluation broadly enough. *Both* description and judgment are essential—in fact, they are the two basic acts of evaluation. Any individual evaluator may attempt to refrain from judging or from collecting the judgments of others. Any individual evaluator may seek only to bring to light the worth of the program. But their evaluations are incomplete. To be fully understood, the educational program must be fully described and fully judged.

## Towards Full Description

The specialist in evaluation seems to be increasing his emphasis on fullness of description. For many years he evaluated primarily by measuring student progress toward academic objectives. These objectives usually were identified with the traditional disciplines, e.g. mathematics, English, and social studies. Achievement tests—standardized or "teacher-made"—were found to be useful in describing the degree to which some curricular objectives are attained by individual students in a particular course. To the early evaluators, and to many others, the countenance of evaluation has been nothing more than the administration and normative interpretation of achievement tests.

In recent years a few evaluators have attempted, in addition, to assess progress of individuals toward certain "inter-disciplinary" and "extracurricular" objectives. In their objectives, emphasis has been given to the integration of behavior within an individual; or to the perception of interrelationships among scholastic disciplines; or to the development of habits, skills, and attitudes which permit the individual to be a craftsman or scholar, in or out of school. For the descriptive evaluation of such outcomes, the Eight-Year Study[2] has served as one model. The proposed National Assessment Program may be another—this statement appeared in one interim report:

> *. . . all committees worked within the following broad definition of 'national assessment:'*
>
> *1. In order to reflect fairly the aims of education in the U. S., the assessment should consider both traditional and modern curricula, and take into account ALL THE ASPIRATIONS schools have for developing attitudes and motivations as well as knowledge and skills . . .* [*Caps added*].[3]

In his paper, "Evaluation for Course Improvement,"[4] Lee Cronbach urged another step: a most generous inclusion of behavioral-science variables in order to examine the possible causes and effects of quality teaching. He proposed that the main objective for evaluation is to uncover durable relationships—those appropriate for guiding future educational programs. To the traditional description of pupil achievement, we add the description of instruction and the description of relationships between them. Like the instructional researcher, the evaluator—as so defined—seeks generalizations about educational practices. Many curriculum project evaluators are adopting this definition of evaluation.

### The Role of Judgment

Description is one thing, judgment is another. Most evaluation specialists have chosen not to judge. But in his recent *Methodology of Evaluation*[5] Michael Scriven has charged evaluators with responsibility for passing upon the merit of an educational practice. (Note that he has urged the evaluator to do what the educator has expected the evaluator to be doing.) Scriven's position is that there is no evaluation until judgment has been passed, and by his reckoning the evaluator is best qualified to judge.

By being well experienced and by becoming well-informed in the case at hand in matters of research and educational practice the evaluator does become at least partially qualified to judge. But is it wise for him to accept this responsibility? Even now when few evaluators expect to judge, educators are reluctant to initiate a formal evaluation. If evaluators were *more* frequently identified with the passing of judgment, with the discrimination among poorer and better programs, and with the awarding of support and censure, their access to data would probably diminish. Evaluators collaborate with other social scientists and behavioral research workers. Those who do not want to judge deplore the acceptance of such responsibility by their associates. They believe that in the eyes of many practitioners, social science and behavioral research will become more suspect than it already is.

Many evaluators feel that they are not capable of perceiving, as they think a judge should, the unidimensional *value* of alternative programs. They anticipate a dilemma such as Curriculum I resulting in three skills and ten understandings and Curriculum II resulting in four skills and eight understandings. They are reluctant to judge that gaining one skill is worth losing two understandings. And, whether through timidity, disinterest, or as a rational choice, the evaluator usually supports "local option," a community's privilege to set its own standards and to be its own judge of the worth of its educational system. He expects that what is good for one community will not necessarily be good for another community, and he does not trust himself to discern what is best for a briefly-known community.

Scriven reminds them that there are precious few who can judge complex programs, and fewer still who will. Different decisions must be made—P.S.S.C. or Harvard Physics?—and they should not be made on trivial criteria, e.g. mere precedent, mention in the popular press, salesman personality, administrative convenience, or pedagogical myth. Who should judge? The answer comes easily to Scriven partly because he expects little interaction between treatment and learner, i.e., what works best for one learner will work best for others, at least within broad categories. He also expects that where the local good is at odds with the common good, the local good can be shown to be detrimental to the common good, to the end that the doctrine of local option is invalidated. According to Scriven the evaluator must judge.

Whether or not evaluation specialists will accept Scriven's challenge remains to be seen. In any case, it is likely that judgments will become an increasing part of the evaluation report. Evaluators will seek out and record the opinions of persons of special qualification. These opinions, though subjective, can be very useful and can be gathered objectively, independent of the solicitor's opinions. A responsibility for processing judgments is much more acceptable to the evaluation specialist than one for rendering judgments himself.

Taylor and Maguire[6] have pointed to five groups having important opinions on education: spokesmen for society at large, subject-matter experts, teachers, parents, and the students themselves. Members of these and other groups are judges who should be heard. Superficial polls, letters to the editor, and other incidental judgments are insufficient. An evaluation of a school program should portray the merit and fault perceived by well-identified groups, systematically gathered and processed. Thus, judgment data and description data are both essential to the evaluation of educational programs.

## Data Matrices

In order to evaluate, an educator will gather together certain data. The data are likely to be from several quite different sources, gathered in several quite different ways. Whether the immediate purpose is description or judgment, three bodies of information should be tapped. In the evaluation report it can be helpful to distinguish between *antecedent*, *transaction*, and *outcome* data.

An antecedent is any condition existing prior to teaching and learning which may relate to outcomes. The status of a student prior to his lesson, e.g. his aptitude, previous experience, interest, and willingness, is a complex antecedent. The programmed-instruction specialist calls some antecedents "entry behaviors." The state accrediting agency emphasizes the investment of community resources. All of these are examples of the antecedents which an evaluator will describe.

Transactions are the countless encounters of students with teacher, student with student, author with reader, parent with counselor—the succession of engagements which comprise the process of education. Examples are the presentation of a film, a class discussion, the working of a homework problem, an explanation on the margin of a term paper, and the administration of a test. Smith and Meux studied such transactions in detail and have provided an 18-category classification system.[7] One very visible emphasis on a particular class of transactions was the National Defense Education Act support of audio-visual media.

Transactions are dynamic whereas antecedents and outcomes are relatively static. The boundaries between them are not clear, e.g. during a transaction we can identify certain outcomes which are feedback antecedents for subsequent learning. These boundaries do not need to be distinct. The categories should be used to stimulate rather than to subdivide our data collection.

Traditionally, most attention in formal evaluation has been given to outcomes—outcomes such as the abilities, achievements, attitudes, and aspirations of students resulting from an educational experience. Outcomes, as a body of information, would include measurements of the impact of instruction on teachers, administrators, counselors, and others. Here too would be data on wear and tear of equipment, effects of the learning environment, cost incurred. Outcomes to be considered in evaluation include not only those that are evident, or even existent, as learning sessions end, but include applications, transfer, and relearning effects which may not be available for measurement until long after. The description of the outcomes of driver training, for example, could well include reports of accident-avoidance over a lifetime. In short, outcomes are the consequences of educating—immediate and long-range, cognitive and conative, personal and community-wide.

Antecedents, transactions, and outcomes, the elements of evaluation statements, are shown in Figure 1 to have a place in both description and judgment. To fill in these matrices the evaluator will collect judgments (e.g. of community prejudice, of problem solving styles, and of teacher personality) as well as descriptions. In Figure 1 it is also indicated that judgmental statements are classified either as general standards of quality or as judgments specific to the given program. Descriptive data are classified as intents and observations. The evaluator can organize his data-gathering to conform to the format shown in Figure 1.

The evaluator can prepare a record of what educators intend, of what ob-

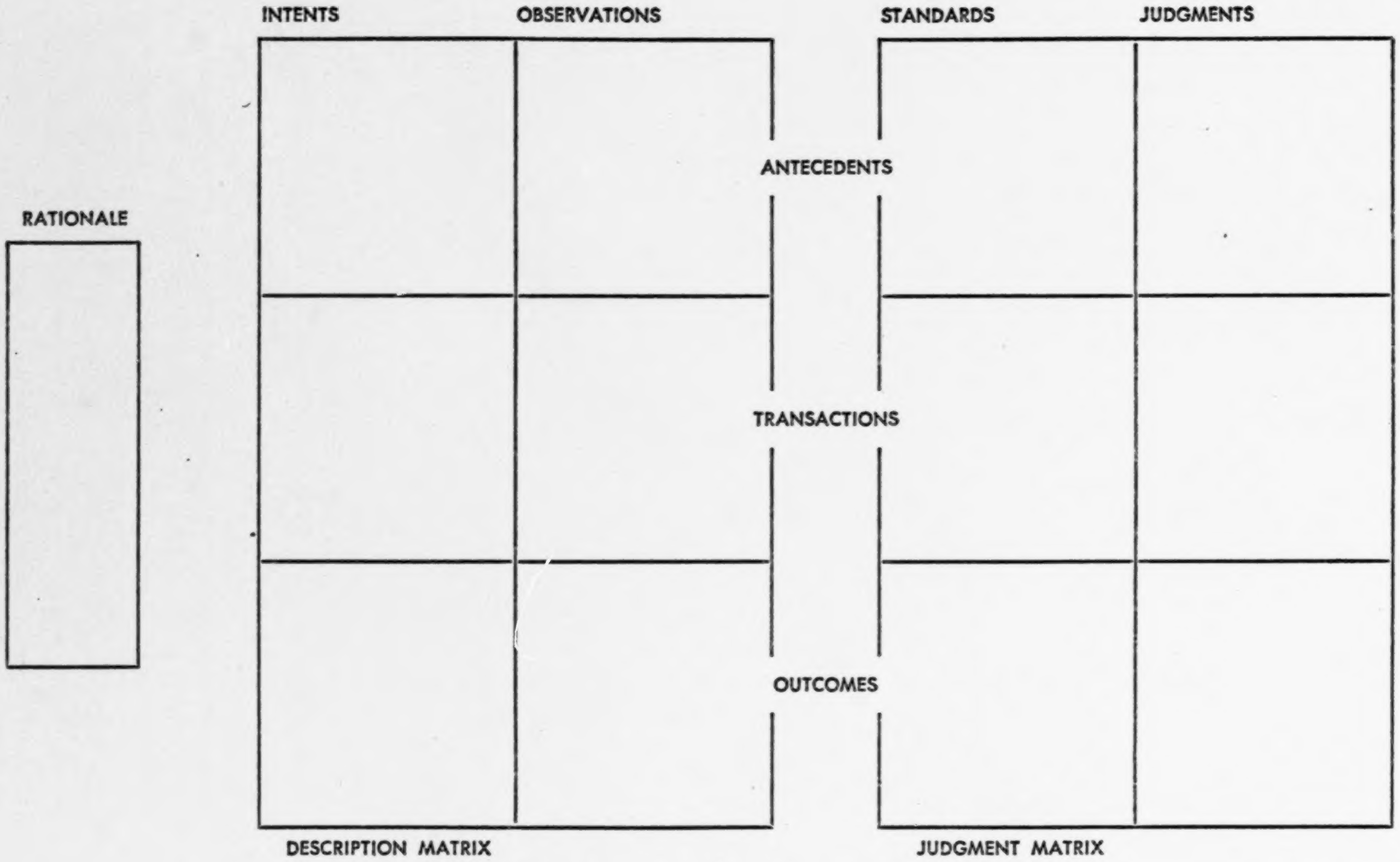

RATIONALE
INTENTS
OBSERVATIONS
STANDARDS
JUDGMENTS
ANTECEDENTS
TRANSACTIONS
OUTCOMES
DESCRIPTION MATRIX
JUDGMENT MATRIX

Figure 1. A layout of statements and data to be collected by the evaluator of an educational program.

servers perceive, of what patrons generally expect, and of what judges value the immediate program to be. The record may treat antecedents, transactions, and outcomes separately within the four classes identified as *Intents, Observations, Standards,* and *Judgments,* as in Figure 1. The following is an illustration of 12 data, one of which could be recorded in each of the 12 cells, starting with an intended antecedent, and moving down each column until an outcome judgment has been indicated.

*Knowing that (1) Chapter XI has been assigned and that he intends (2) to lecture on the topic Wednesday, a professor indicates (3) what the students should be able to do by Friday, partly by writing a quiz on the topic. He observes that (4) some students were absent on Wednesday, that (5) he did not quite complete the lecture because of a lengthy discussion and that (6) on the quiz only about 2/3 of the class seemed to understand a certain major concept. In general, he expects (7) some absences but that the work will be made up by quiz-time; he expects (8) his lectures to be clear enough for perhaps 90 percent of a class to follow him without difficulty; and he knows that (9) his colleagues expect only about one student in ten to understand thoroughly each major concept in such lessons as these. By his own judgment (10) the reading assignment was not a sufficient background for his lecture; the students commented that (11) the lecture was provocative; and the graduate assistant who read the quiz papers said that (12) a discouragingly large number of students seemed to confuse one major concept for another.*

Evaluators and educators do not expect data to be recorded in such detail, even in the distant future. My purpose here was to give twelve examples of data that could be handled by separate cells in the matrices. Next I would like to consider the description data matrix in detail.

### Goals and Intents

For many years instructional technologists, test specialists, and others have pleaded for more explicit statement of educational goals. I consider "goals," "objectives," and "intents" to be synonomous. I use the category title *Intents* because many educators now equate "goals" and "objectives" with "intended student outcomes." In this paper Intents includes the planned-for environmental conditions, the planned-for demonstrations, the planned-for coverage of certain subject matter, etc., as well as the planned-for student behavior. To be included in this three-cell column are effects which are desired, those which are hoped for, those which are anticipated, and even those which are feared. This class of data includes goals and plans that others have, especially the students. (It should be noted that it is not the educator's privilege to rule out the study of a variable by saying, "that is not one of our objectives." The evaluator should include both the variable and the negation.) The resulting collection of *Intents* is a priority listing of all that may happen.

The fact that many educators now equate "goals" with "intended student outcomes" is to the credit of the behaviorists, particularly the advocates of programmed instruction. They have brought about a small reform in teaching by emphasizing those specific classroom acts and work exercises which contribute to the refinement of student responses. The A.A.A.S. Science Project, for example, has been successful in developing its curriculum around behavioristic goals.[8] Some curriculum-innovation projects, however, have found the emphasis on behavioral outcomes an obstacle to creative teaching.[9] The educational evaluator should not list goals only in terms of anticipated student behavior. To *evaluate* an educational program, we must examine what teaching, as well as what learning, is intended. (Many antecedent conditions and teaching transactions can be worded behavioristically, if desired.) How intentions are worded is not a criterion for inclusion. Intents can be the global goals of the Educational Policies Commission or the detailed goals of the programmer.[10] Taxonomic, mechanistic, humanistic, even scriptural—any mixture of goal statements are acceptable as part of the evaluation picture.

Many a contemporary evaluator expects trouble when he sets out to record the educator's objectives. Early in the work he urged the educator to declare his objectives so that outcome-testing devices could be built. He finds the educator either reluctant or unable to verbalize objectives. With diligence, if not with pleasure, the evaluator assists with what he presumes to be the educator's job: writing behavioral goals. His presumption is wrong. As Scriven has said, the responsibility for describing curricular objectives is the responsibility of the evaluator. He is the one who is experienced with the language of behaviors, traits, and habits. Just as it is his responsibility to transform the behaviors of a teacher and the responses of a student into data, it is his responsibility to transform the intentions and expectations of an educator into "data." It is necessary for him to continue to ask the educator for statements of intent. He should augment the replies by asking, "Is this another way of saying it?" or "Is this an instance?" It is not wrong for an evaluator to teach a willing educator about behavioral objectives—they may facilitate the work. It is wrong for him to insist that every educator should use them.

Obtaining authentic statements of intent is a new challenge for the evaluator. The methodology remains to be developed. Let us now shift attention to the second column of the data cells.

### Observational Choice

Most of the descriptive data cited early in the previous section are classified as *Observations*. In Figure 1 when he described surroundings and events and the subsequent consequences, the evaluator* is telling of his Observations. Sometimes the evaluator observes these characteristics in a direct and personal way. Sometimes he uses instruments. His instruments include inventory schedules, biographical data sheets, interview routines, check lists, opinionnaires, and all kinds of psychometric tests. The experienced evaluator gives special attention to the measurement of student outcomes, but he does not fail to observe the other outcomes, nor the antecedent conditions and instructional transactions.

---

* Here and elsewhere in this paper, for simplicity of presentation, the evaluator and the educator are referred to as two different persons. The educator will often be his own evaluator or a member of the evaluation team.

Many educators fear that the outside evaluator will not be attentive to the characteristics that the school staff has deemed most important. This sometimes does happen, but evaluators often pay *too much* attention to what they have been urged to look at, and too little attention to other facets. In the matter of selection of variables for evaluation, the evaluator must make a subjective decision. Obviously, he must limit the elements to be studied. He cannot look at all of them. The ones he rules out will be those that he assumes would not contribute to an understanding of the educational activity. He should give primary attention to the variables specifically indicated by the educator's objectives, but he must designate additional variables to be observed. He must search for unwanted side effects and incidental gains. The selection of measuring techniques is an obvious responsibility, but the choice of characteristics to be observed is an equally important and unique contribution of the evaluator.

An evaluation is not complete without a statement of the rationale of the program. It needs to be considered separately, as indicated in Figure 1. Every program has its rationale, though often it is only implicit. The rationale indicates the philosophic background and basic purposes of the program. Its importance to evaluation has been indicated by Berlak.[11] The rationale should provide one basis for evaluating Intents. The evaluator asks himself or other judges whether the plan developed by the educator constitutes a logical step in the implementation of the basic purposes. The rationale also is of value in choosing the reference groups, e.g. merchants, mathematicians, and mathematics educators, which later are to pass judgment on various aspects of the program.

A statement of rationale may be difficult to obtain. Many an effective instructor is less than effective at presenting an educational rationale. If pressed, he may only succeed in saying something the listener wanted said. It is important that the rationale be in his language, a language he is the master of. Suggestions by the evaluator may be an obstacle, becoming accepted because they are attractive rather than because they designate the grounds for what the educator is trying to do.

The judgment matrix needs further explanation, but I am postponing that until after a consideration of the bases for processing descriptive data.

### Contingency and Congruence

For any one educational program there are two principal ways of processing descriptive evaluation data: finding the contingencies among antecedents, transactions, and outcomes and finding the congruence between Intents and Observations. The processing of judgments follows a different model. The first two main columns of the data matrix in Figure 1 contain the descriptive data. The format for processing these data is represented in Figure 2.

The data for a curriculum are *congruent* if what was intended actually happens. To be fully congruent the intended antecedents, transactions, and outcomes would have to come to pass. (This seldom happens—and often should

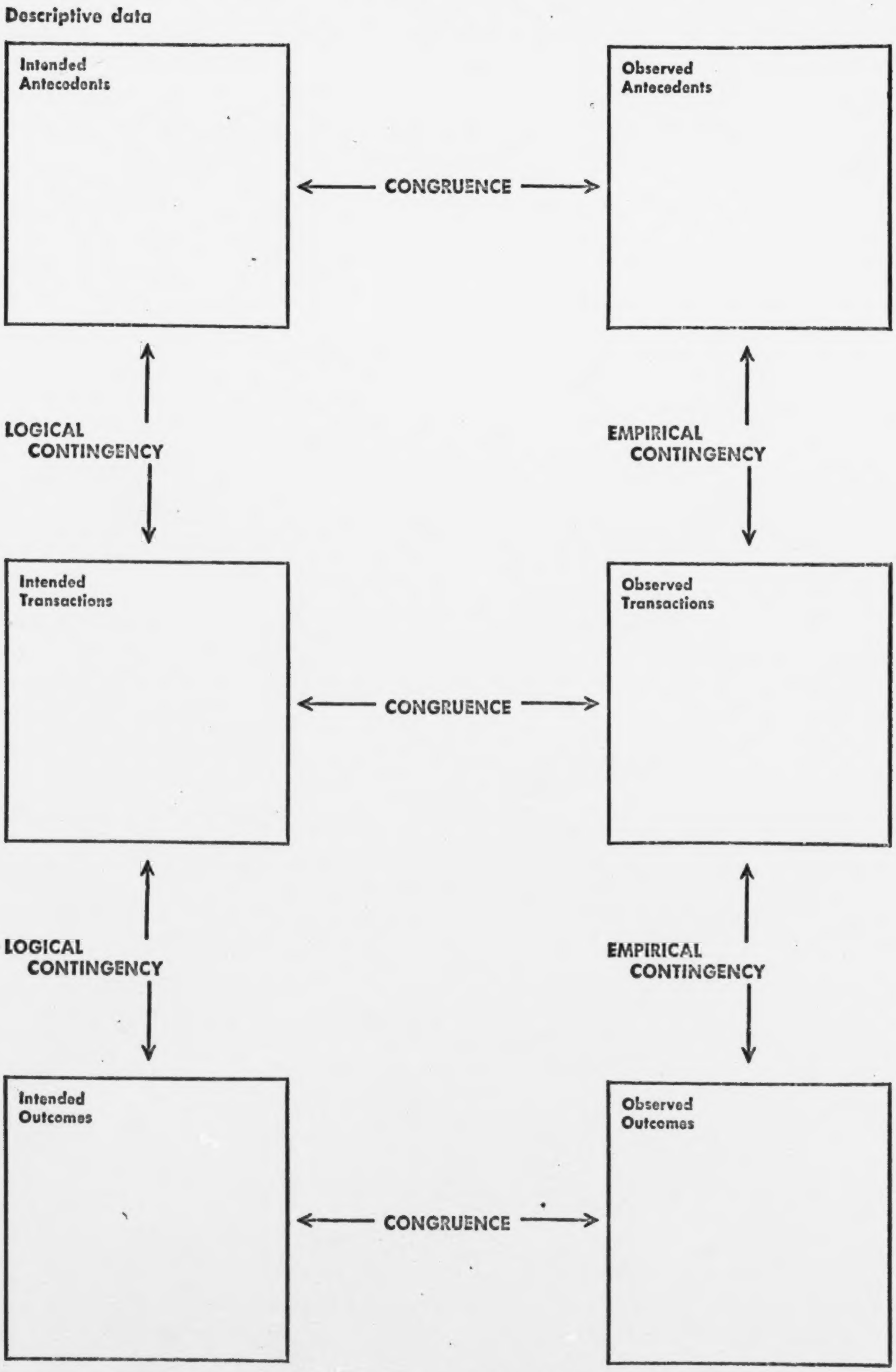

Descriptive data
Intended Antecedents
CONGRUENCE
Observed Antecedents
LOGICAL CONTINGENCY
EMPIRICAL CONTINGENCY
Intended Transactions
CONGRUENCE
Observed Transactions
LOGICAL CONTINGENCY
EMPIRICAL CONTINGENCY
Intended Outcomes
CONGRUENCE
Observed Outcomes

Figure 2. A representation of the processing of descriptive data.

not.) Within one row of the data matrix the evaluator should be able to compare the cells containing Intents and Observations, to note the discrepancies, and to describe the amount of congruence for that row. (Congruence of outcomes has been emphasized in the evaluation model proposed by Taylor and Maguire.) Congruence does not indicate that outcomes are reliable or valid, but that what was intended did occur.

Just as the Gestaltist found more to the whole than the sum of its parts, the evaluator studying variables from any two of the three cells in a column of the data matrix finds more to describe than the variables themselves. The relationships or *contingencies* among the variables deserve additional attention. In the sense that evaluation is the search for relationships that permit the improvement of education, the evaluator's task is one of identifying outcomes that are contingent upon particular antecedent conditions and instructional transactions.

Lesson planning and curriculum revision through the years has been built upon faith in certain contingencies. Day to day, the master teacher arranges his presentation and selects his input materials to fit his instructional goals. For him the contingencies, in the main, are logical, intuitive, and supported by a history of satisfactions and endorsements. Even the master teacher and certainly less-experienced teachers need to bring their intuited contingencies under the scrutiny of appropriate juries.

As a first step in evaluation it is important just to record them. A film on floodwaters may be scheduled (intended transaction) to expose students to a background to conservation legislation (intended outcome). Of those who know both subject matter and pedagogy, we ask, "Is there a logical connection between this event and this purpose?" If so, a logical contingency exists between these two Intents. The record should show it.

Whenever Intents are evaluated the contingency criterion is one of logic. To test the logic of an educational contingency the evaluators rely on previous experience, perhaps on research experience, with similar observables. No immediate observation of these variables, however, is necessary to test the strength of the contingencies among Intents.

Evaluation of Observation contingencies depends on empirical evidence. To say, "this arithmetic class progressed rapidly because the teacher was somewhat but not too sophisticated in mathematics" demands empirical data, either from within the evaluation or from the research literature.[12] The usual evaluation of a single program will not alone provide the data necessary for contingency statements. Here too, then, previous experience with similar observables is a basic qualification of the evaluator.

The contingencies and congruences identified by evaluators are subject to judgment by experts and participants just as more unitary descriptive data are. The importance of non-congruence will vary with different viewpoints. The school superintendent and the school counselor may disagree as to the importance of a cancellation of the scheduled lessons on sex hygiene in the health class. As an example of judging contingencies, the degree to which teacher morale is contingent on the length of the school day may be deemed cause enough to abandon an early morning class by one judge and not another. Perceptions of importance of congruence and contingency deserve the evaluator's careful attention.

### Standards and Judgments

There is a general agreement that the goal of education is excellence—but how schools and students should excell, and at what sacrifice, will always be debated. Whether goals are local or national, the measurement of excellence requires explicit rather than implicit standards.

Today's educational programs are not subjected to "standard-oriented" evaluation. This is not to say that schools lack in aspiration or accomplishment. It is to say that standards—benchmarks of performance having widespread reference value—are not in common use. Schools across the nation may use the same evaluation checklist** but the interpretations of the checklisted data are couched in inexplicit, personal terms. Even in an informal way, no school can evaluate the impact of its program without knowledge of what other schools are doing in pursuit of similar objectives. Unfortunately, many educators are loathe to accumulate that knowledge systematically.[13, 14]

There is little knowledge anywhere today of the quality of a student's education. School grades are based on the private criteria and standards of the individual teacher. Most "standardized" tests scores tell where an examinee performing "psychometrically useful" tasks stands with regard to a reference group, rather than the level of competence at which he performs essential scholastic tasks. Although most teachers are competent to teach their subject matter and to spot learning difficulties, few have the ability to *describe* a student's command over his intellectual environment. Neither school grades nor standardized test scores nor the candid opinions of teachers are very informative as to the excellence of students.

Even when measurements are effectively interpreted, evaluation is complicated by a multiplicity of standards. Standards vary from student to student, from instructor to instructor, and from reference group to reference group. This is not wrong. In a healthy society, different parties have different standards. Part of the responsibility of evaluation is to make known which standards are held by whom.

It was implied much earlier that it is reasonable to expect change in an educator's *Intents* over a period of time. This is to say that he will change both his criteria and his standards during instruction. While a curriculum is being developed and disseminated, even the major classes of criteria vary. In their analysis of nationwide assimilation of new educational programs, Clark and Guba[15] identified eight stages of change through which new pro-

---

** One contemporary checklist is *Evaluative Criteria,* a document published by the National Study of Secondary School Evaluation (1960). It is a commendably thorough list of antecedents and possible transactions, organized mostly by subject-matter offerings. Surely it is valuable as a checklist, identifying neglected areas. Its great value may be a catalyst, hastening the maturity of a developing curriculum. However, it can be of only limited value in *evaluating,* for it guides neither the measurement nor the interpretation of measurement. By intent, it deals with criteria (what variables to consider) and leaves the matter of standards (what ratings to consider as meritorious) to the conjecture of the individual observer.

grams go. For each stage they identified special criteria (each with its own standards) on which the program should be evaluated before it advances to another stage. Each of their criteria deserves elaboration, but here it is merely noted that there are quite different criteria at each successive curriculum-development stage.

Informal evaluation tends to leave criteria unspecified. Formal evaluation is more specific. But it seems the more careful the evaluation, the fewer the criteria; and the more carefully the criteria are specified, the less the concern given to standards of acceptability. It is a great misfortune that the best trained evaluators have been looking at education with a microscope rather than with a panoramic view finder.

There is no clear picture of what any school or any curriculum project is accomplishing today partly because the methodology of processing judgments is inadequate. What little formal evaluation there is is attentive to too few criteria, overly tolerant of implicit standards, and ignores the advantage of relative comparisons. More needs to be said about relative and absolute standards.

### Comparing and Judging

There are two bases of judging the characteristics of a program, (1) with respect to absolute standards as reflected by personal judgments and (2) with respect to relative standards as reflected by characteristics of alternate programs. One can evaluate SMSG mathematics with respect to opinions of what a mathematics curriculum should be or with regard to what other mathematics curricula are. The evaluator's comparisons and judgments are symbolized in Figure 3. The upper left matrix represents the data matrix from Figure 2. At the upper right are sets of standards by which a program can be judged in an absolute sense. There are multiple sets because there may be numerous reference groups or points of view. The several matrices at the lower left represent several alternate programs to which the one being evaluated can be compared.

Each set of absolute standards, if formalized, would indicate acceptable and meritorious levels for antecedents, transactions, and outcomes. So far I have been talking about setting standards, not about judging. Before making a judgment the evaluator determines whether or not each standard is met. Unavailable standards must be estimated. The judging act itself is deciding which set of standards to heed. More precisely, judging is assigning a weight, an importance, to each set of standards. Rational judgment in educational evaluation is a decision as to how much to pay attention to the standards of each reference group (point of view) in deciding whether or not to take some administrative action.‡

Relative comparison is accomplished in similar fashion except that the standards are taken from descriptions of other programs. It is hardly a judgmental matter to determine whether one program betters another with

---

‡ Deciding which variables to study and deciding which standards to employ are two essentially subjective commitments in evaluation. Other acts are capable of objective treatment; only these two are beyond the reach of social science methodology.

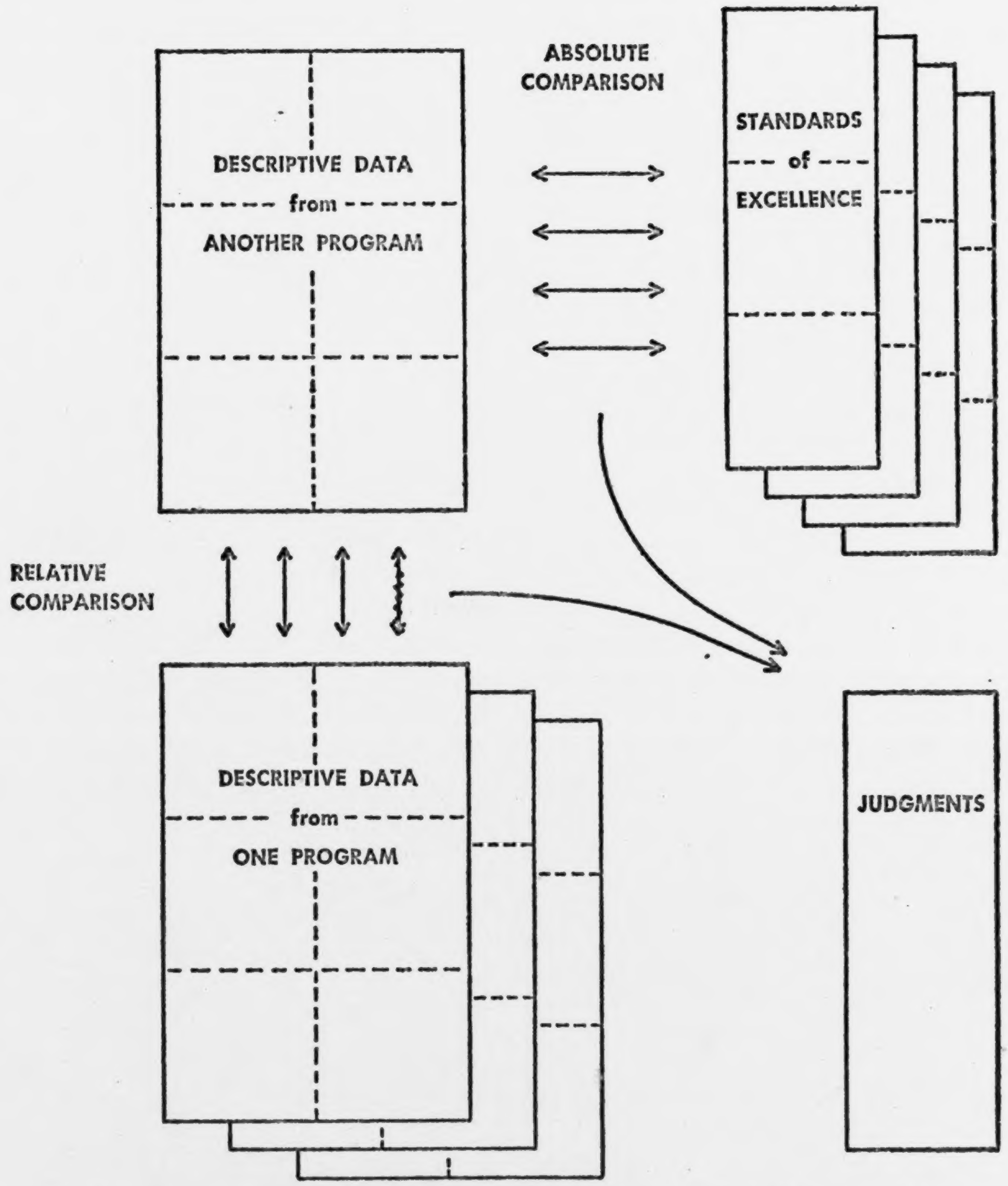

ABSOLUTE COMPARISON
DESCRIPTIVE DATA from ANOTHER PROGRAM
STANDARDS of EXCELLENCE
RELATIVE COMPARISON
DESCRIPTIVE DATA from ONE PROGRAM
JUDGMENTS

Figure 3. A representation of the process of judging the merit of an educational program.

regard to a single characteristic, but there are many characteristics and the characteristics are not equally important. The evaluator selects which characteristics to attend to and which reference programs to compare to. From relative judgment of a program, as well as from absolute judgment we can obtain an overall or composite rating of merit (perhaps with certain qualifying statements), a rating to be used in making an educational decision. From this final act of judgment a recommendation can be composed.

### Absolute and Relative Evaluation

As to which kind of evaluation—absolute or relative—to encourage, Scriven and Cronbach have disagreed. Cronbach[4] suggests that generalizations to the local-school situation from curriculum-comparing studies are sufficiently hazardous (even when the studies are massive, well-designed, and properly controlled) to make them poor research investments. Moreover, the difference in purpose of the programs being compared is likely to be sufficiently great to render uninterpretable any outcome other than across-the-board superiority of one of them. Expecting that rarely, Cronbach urges fewer comparisons, more intensive process studies, and more curriculum "case studies" with extensive measurement and thorough description. Scriven, on the other hand, indicates that what the educator wants to know is whether or not one program is better than another, and that the best way to answer his question is by direct comparison. He points to the difficulty of describing the outcomes of complex learning in explicit terms and with respect to absolute standards, and to the ease of observing relative outcomes from two programs. Whether or not Scriven's prescription is satisfying will probably depend on the client. An educator faced with an adoption decision is more likely to be satisfied, the curriculum innovator and instructional technologist less likely.

One of the major distinctions in evaluation is that which Scriven identifies as *formative* versus *summative* evaluation. His use of the terms relates primarily to the stage of development of curricular material. If material is not yet ready for distribution to classroom teachers, then its evaluation is formative; otherwise it is summative. It is probably more useful to distinguish between evaluation oriented to developer-author-publisher criteria and standards and evaluation oriented to consumer-administrator-teacher criteria and standards. The formative-summative distinction could be so defined, and I will use the terms in that way. The faculty committee facing an adoption choice asks, "Which is best? Which will do the job best?" The course developer, following Cronbach's advice, asks, "How can we teach it better?" (Note that neither are now concerned about the individual student differences.) The evaluator looks at different data and invokes different standards to answer these questions.

The evaluator who assumes responsibility for summative evaluation—rather than formative evaluation—accepts the responsibility of informing consumers as to the merit of the program. The judgments of Figure 3 are his target. It is likely that he will attempt to describe the school situations in which the procedures or materials may be used. He may see his task as one of indicating the goodness-of-fit of an available curriculum to an existing school program. He must learn whether or not the intended antecedents, transactions, and outcomes for the curriculum are consistent with the resources, standards, and goals of the school. This may require as much attention to the school as to the new curiculum.

The formative evaluator, on the other hand, is more interested in the contingencies indicated in Figure 2. He will look for covariations within the evaluation study, and across studies, as a basis for guiding the development of present or future programs.

For major evaluation activities it is obvious that an individual evaluator will not have the many competencies required. A team of social scientists is needed for many assignments. It is reasonable to suppose that such teams will include specialists in instructional technology, specialists in psychometric testing and scaling, specialists in research design and analysis, and specialists in dissemination of information. Curricular innovation is sure to have deep and widespread effect on our society, and we may include the social anthropologist on some evaluation teams. The economist and philosopher have something to offer. Experts will be needed for the study of values, population surveys, and content-oriented data-reducation techniques.

The educator who has looked disconsolate when scheduled for evaluation will look aghast at the prospect of a team of evaluators invading his school. How can these evaluators observe or describe the natural state of education when their very presence influences that state? His concern is justified. Measurement activity—just the presence of evaluators—does have a reactive effect on education, sometimes beneficial and sometimes not—but in either case contributing to the atypicality of the sessions. There are specialists, however, who anticipate that evaluation will one day be so skilled that it properly will be considered "unobtrusive measurement."[16]

In conclusion I would remind the reader that one of the largest investments being made in U. S. education today is in the development of new programs. School officials cannot yet revise a curriculum on rational grounds, and the needed evaluation is not under way. What is to be gained from the enormous effort of the innovators of the 1960's if in the 1970's there are no evaluation records? Both the new innovator and the new teacher need to know. Folklore is not a sufficient repository. In our data banks we should document the causes and effects, the congruence of intent and accomplishment, and the panorama of judgments of those concerned. Such records should be kept to promote educational action, not obstruct it. The countenance of evaluation should be one of data gathering that leads to decision-making, not to trouble-making.

Educators should be making their own evaluations more deliberate, more formal. Those who will—whether in their classrooms or on national panels—can hope to clarify their responsibility by answering each of the following questions: (1) Is this evaluation to be primarily descriptive, primarily judgmental, or both descriptive and judgmental? (2) Is this evaluation to emphasize the antecedent conditions, the transactions, or the outcomes alone, or a combination of these, or their functional contingencies? (3) Is this evaluation to indicate the congruence between what is intended and what occurs? (4) Is this evaluation to be undertaken within a single program or as a comparison between two or more curricular programs? (5) Is this evaluation intended more to further the development of curricula or to help choose among available curricula? With these questions answered, the restrictive effects of incomplete guidelines and inappropriate countenances are more easily avoided.

## References

[1] American Council on Education. *Educational Measurement.* E. F. Lindquist (Ed.). Washington, D. C., 1951.

[2] Smith, E. R. and Tyler, Ralph W., *Appraising and Recording Student Progress.* New York: Harper and Row, 1942.

[3] Educational Testing Service. "A Long, Hot Summer of Committee Work on National Assessment of Education," *ETS Developments,* Vol. XIII, November, 1965.

[4] Cronbach, Lee. "Evaluation for Course Improvement," *Teachers College Record,* 64, 1963, pp. 672-683.

[5] Scriven, Michael. "The Methodology of Evaluation," *AERA Monograph Series on Curriculum Evaluation,* No. 1. Chicago: Rand McNally, 1967, pp. 39-89.

[6] Taylor, Peter A. and Maguire, Thomas O., "A Theoretical Evaluation Model," *The Manitoba Journal of Educational Research,* 1, 1966, pp. 12-17.

[7] Smith, B. Othanel and Meux, M. O., *A Study of the Logic of Teaching.* Urbana: Bureau of Educational Research, University of Illinois. No date.

[8] Gagné, Robert M., "Elementary Science: A New Scheme of Instruction," *Science,* Vol. 151, No. 3706, pp. 49-53.

[9] Atkin, J. M., "Some Evaluation Problems in a Course Content Improvement Project," *Journal of Research in Science Teaching,* 1, 1963, 129-132.

[10] Mager, R. F., *Preparing Objectives For Programmed Instruction.* San Francisco: Fearon Publishers, 1962.

[11] Berlak, Harold. Comments recorded in *Concepts and Structure in the New Social Science Curricula.* Irving Morrissett (Ed.). Lafayette, Indiana: Social Science Education Consortium, Purdue University, 1966, pp. 88-89.

[12] See Bassham, H., "Teacher Understanding and Pupil Efficiency in Mathematics: A Study of Relationship," *Arithmetic Teacher,* 9: 1962, pp. 383-387.

[13] Hand, Harold C., "National Assessment Viewed as the Camel's Nose," *Phi Delta Kappan,* 47, September, 1965, 8-12.

[14] Tyler, Ralph W., "Assessing the Progress of Education," *Phi Delta Kappan,* 47, September, 1965, 13-16.

[15] Clark, David L. and Guba, Egon G., "An Examination of Potential Change Roles in Education." Columbus: The Ohio State University, 1965. (Multilith.)

[16] Webb, Eugene J.; Campbell, Donald T., Schwartz, Richard D., and Sechrist, Lee. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally, 1966.

# GENERALIZABILITY OF PROGRAM EVALUATION:

## *THE NEED FOR LIMITS*

Educational Products Report, Febr., 1969.

*By Robert E. Stake*

What should a school research director or Title III project director be doing about program evaluation? This is a tough question to answer. The textbooks on testing and statistics and research design do not tell us, and to my knowledge, there are still no handbooks of evaluation or sets of guidelines for new projects that satisfactorily provide the answers.

This is also a difficult question because programs vary so radically from place to place and because different audiences—school boards, parents, state departments, Congress, as well as members of the staff—have different notions as to what information they want to find in an evaluation report. One of the biggest complications is that consultants on evaluation are giving advice about what information to gather as if the evaluation were a traditional research project.

The contrast between educational program evaluation and educational research must be emphasized. I am sure that in many school situations, the research director and the program evaluator should do quite different things. Part of the answer to what the evaluator should do depends on how much and in what directions the findings are expected to generalize, to be relevant to programs other than the one observed.

We expect a research project to provide us with generalizable findings. An educational researcher goes to graduate school to learn some of the methods of the social sciences and the behavioral sciences. If he applies these scientific methods, he can claim, at some level of confidence, that what has happened in his study will happen in a similar way elsewhere. The school research worker is expected—at least by his instructors in research methods—to be a scientist, to seek generalizations about why school programs work or do not work.

The question I raise here is, "Should evaluators be scientific?" Of course, evaluation should be logical, empirical, and objective, but how about scientific? The answer is: "yes" on some occasions, on others "no." There are higher species of evaluation that are entirely within the scientific process and there are lower species—healthy, beneficial species, but lower species of evaluation—that are outside the scientific process. For many an audience, for many an evaluation job, the program evaluator should choose a lower form of evaluation rather than a scientific-research form of evaluation.

The two fields of inquiry called "evaluation" and "scientific research" overlap, but neither envelopes the other. All evaluation deals explicitly with the worth of something. Only a few research studies do. It is this latter distinction—inquiry for generalization versus inquiry for specification--that I am emphasizing.

In speaking of higher and lower forms of evaluation, I am referring to the generalization that results from them. A higher form of evaluation permits generalization in many directions. Findings are expected to hold over different school buildings, for different types of teachers, across communities, and over replications. A lower form of evaluation yields conclusions limited to a specific setting, perhaps to a particular kind of classroom and to a particular kind of student, sometimes specific even to a given occasion. We can place *little* confidence in extensions of these findings to other settings or later occasions. Extent of generalizability is the major difference as I see it between high and low forms of evaluation.

The distinction between instructional research and educational program evaluation can also be made on the grounds that, unlike the researcher, the program evaluator has a primary concern for a designated program. His might be a tiny program, such as a unit on homeostasis, or it might be a gigantic program, such as Headstart in the United States. The program is specified, the setting is specified, the people are specified.

When we "evaluate" a particular remedial reading program, we usually do not look at it as representative of others, but for its value as it is. Several programs can be evaluated in the same study, but the evaluation study focuses directly on just those that are named, rather than on all such programs.

In an evaluative study, some one or more programs are the target programs. To put it in the language of analysis of variance, provided by H. A. Scheffé, the treatments (programs) are a "fixed effect." There is no effort to sample from the population of programs. There is no built-in "scientific" basis for interference from the programs that are evaluated to those which are not. In this one dimension, all evaluation study findings are restricted in generalizability. In contrast, instructional research often has fixed treatments, but it usually is designed to generalize over what we consider to be "programs."

There are other "main effects," of course, besides treatments (programs). There are students, teachers, classrooms, communities, and many more. The evaluation may be designed to generalize over some effects and to be fixed on others. The study may be designed to investigate how the program works with various kinds of students, but only within a single community. The research study may be designed with the same constraints. The fact that researchers are trained to work toward broad generalizations and tend to be curious about broad generalizations is neither a suitable basis for designing an evaluation study to be highly generalizable or for designing it to be highly localized. The degree to which the evaluation study should generalize should be based upon the interests of the people who are waiting for the findings. We strive to make research generalizable; we can make evaluation generalizable, but we are not obligated to do so.

In evaluation circles, the terms *formative* and *summative* are heard more and more frequently. These terms have a dramatic effect, distinguishing between what is done *during development* and what is done *when development is finished.* For the purpose of choosing an evaluation strategy, I find this a trivial distinction. For most educational programs—correspondence courses or Montessori programs—development never ends. For a learner, there is a beginning and an end, but for the teacher, the programs are ongoing, ever evolving. What is important is that there are differences between what the "program people" want to know about their program and what "outsiders" want to know. We can make a non-trivial distinction between formative evaluation for the program developer who is planning ahead and trying to choose the best ingredients, and summative evaluation for anyone who is looking at the program, past or present, and who is trying to find out what it is and what it does.

Let us examine this distinction further. Think of insiders as program developers and outsiders as program consumers. Differences in curiosity between developers and consumers often relate to the specificity of the *conditions of use.* The developer often wants the program to be useful everywhere, in San Francisco *and* Harlan County, Kentucky, and places beyond. He wants to appeal to everyone, to the PTA *and* the U.S. Office of Education and groups he has never heard of. He wants to succeed with children of many backgrounds, many interests, and many aspirations. The product developer *may* have certain target populations in mind, but otherwise he wants the program to be used in many diverse settings. The consumer has a more specific setting in which he wants the product to work. In that setting, there will be some uniqueness of children, parents, teachers, and other groups—they constitute relatively specific populations. The developer needs to design his study to provide a sound basis for generalization to other populations.

Let us consider the extreme case. Suppose you are curious about a once-and-once-only program. You are not interested in a rerun. It could be the evaluation of what you have taught your youngest child, training of the astronauts, or of the Centennial Year lectureships at the University of Illinois. It can be the evaluation of anything. The important thing, in this case, is that a full and exact description is planned for. No generalization is desired; there is no interest in a repeat performance. This is the lowest species of evaluation. It is ascientific. Logical, empirical, and objective it may be, but ascientific. Nevertheless, if it provides the needed descriptive data and judgments of merit, this is a respectable evaluation study for the educator or Congressman who wants to know *what* is happening, not why.

Other summative evaluations *do* invite generalization. Consider the question, "*Who else* could use this program?" Expecting generalization over a population of users, the evaluator will treat certain user characteristics as random effects. He may or may not be hoping to find differential success with different users or in different settings, but he is expecting to generalize and he is concerned about the limits. When will this program work? When won't it?

As long as the program components remain fixed, not subject to modification or reorganization, our efforts to evaluate them can be called *summative* evaluation. In contrast, *formative* evaluation more nearly approximates conventional instructional research. Considering a new syllabus, does a different arrangement of the topics result in better retention of the concepts? Considering a course in geography, should field trips be used? Informative evaluation program *components* may be treated as random *or* fixed effects. Conditions-of-use may or may not have important interaction effects. Here the program is specified, but its components are subject to change. How do program characteristics affect educational outcomes? The formative-evaluation study seeks generalizations about how to create a specific instructional treatment that may become a feature in many educational programs.

Now, the purpose of all this is to offer some handles for grasping the evaluation responsibility. I have found the concepts of formative and summative evaluation useful in taking the first steps toward writing an evaluation plan. They help me decide on who is to benefit from this study, what questions will be asked, what variables will be measured, what generalizations will be sought.

Sometimes an evaluation will broadcast to a wide audience of educators and researchers. The major question will seek out—as Tom Hastings put it—the "whys of the outcomes." The findings should generalize across populations of children, of schoolmen, of committees, across many conditions of use. These educators are not asking, "What is happening?" but "Why does it happen?" Their concern calls for formative evaluation—and for them the evaluator needs to design a study which accounts for outcome variance in terms of product differences.

At other times in other places, the purpose of evaluation is to aid educators using a specific product in a specific setting. The major question seeks out *what* it is that is happening. Findings are not for the benefit of researchers, nor for educators elsewhere. For themselves they are asking, "What is at work?" and "What is it accomplishing?" They need summative evaluation —and for them the evaluator should design a study which describes the purposes, the plans, the background conditions, the transactions, the outcomes and which collects judgments of merit and shortcomings about all these facts.

"Formative-summative." "The specific and the general." "The what and the why." To these distinctions my colleagues usually react by saying, "These are nominal, arbitrary, pedantic distinctions. Every evaluation ought to answer both the practical question, 'What is it?' and the scientific question, 'What makes it go?'" But it seems to me that we cannot possibly answer both questions at the same time. One question denies the other.

Instruction is a complex process. In the field, in the classroom, even using the best of the anthropologist's skills, we cannot detect the ingredient that is present in optimum proportion; we cannot tell what is active and what is inert and what is catalytic; we cannot tell what is *causing* what. The natural variation and covariation is too infrequent and capricious to be a dependable basis for generalization.

To find out *why*, we must invoke scientific methods. But as soon as we exercise a reasonable degree of experimental control, as soon as we provoke some variability in the program and hold other aspects constant, the product is altered. Many an educator finds the program being researched *no* longer the program he wanted to know about.

It seems to me that like the physicists, we have our own *uncertainty principle* in educational evaluation. You cannot simultaneously know *what it is* and *why it is*. There are two approaches. We have a fundamental choice: to be scientific, to generalize, to evaluate to find out *why;* or to be descriptive, to be delimited, and to evaluate to find out *what*.

# EDUCATIONAL OBJECTIVES

1. In any teaching a great number of objectives are simultaneously pursued. High-priority, immediate objectives should usually be apparent to teacher and learner alike. Occasionally, either will do better *without* being aware of them. High-quality education is often accomplished by educators having but a partial awareness of the objectives. Sometimes it will increase teaching-learning effectiveness to make participants more aware of objectives; sometimes it will not.

2. With all who share the responsibility of educating lies the responsibility for stating objectives, arranging environments, providing stimulation, evoking responses, and evaluating those responses. But each author and teacher does not share equally in those responsibilities. Time and talent are not available in limitless abundance. Each educator's assignment should capitalize on what he can do best. Few classroom teachers are skilled in stating objectives. Most are more highly skilled in adapting teaching to immediate circumstances, motivating students, and appraising responses. In the interests of effectiveness, seldom should they be required to formulate behavioral specifications.

3. There are more objectives to pursue than we can pursue. Time and resources restrict us. We assign priorities to our goals in a highly informal way. Even this informal priority list is not always the critical determinant of the daily lesson or the minute-by-minute dialogue. Some moments are ripe for teaching toward an unplanned objective. A sound educational system is one which provides for occasional reassignment of immediate objectives to take advantage of the special opportunities that occur.

4. The development of a new curricular program or set of instructional materials often proceeds better by successive approximations than by linear programming. With successive approximations, major attention is given to getting an enterprise in operation, even though the initial runs are crude and faulty, so that corrections can be based on experience. With linear programming, major attention is given to planning, precise specification, and symbolic representation so that corrections can be based on logical analysis. Advice on curriculum planning should be oriented to the experiential and logical skills already developed in the developers or that can be readily obtained by them.

5. For creating lists of objectives, the technology of education should have some methods that rely on behavioral specification and symbolic delimitation and other methods that rely on illustrative examples and inferable definitions. We need methods by which educators and others can endorse, reject, or revise statements of objectives. Two colossal problems lie before us: how to *translate* global objectives into specific behavioral objectives and how to *derive* appropriate teaching tactics.

6. Our curriculum-development projects and our evaluation studies seldom reach a satisfactory specification by asking educators to state their objectives. Educator's global objectives give little guidance to teaching and evaluation. Their specific objectives ignore vast concerns that they have. In our present state the derivation of the specific from the general is some form of intuitive magic. Luckily it often works pretty well. We need to understand it, to simulate it, not necessarily to replace it.

Stake
3/68

# MULTIPLE CRITERION MEASURES FOR EVALUATION OF SCHOOL PROGRAMS*

Newton S. Metfessel and William B. Michael
University of Southern California

## I. Indicators of Status or Change in Cognitive and Affective Behaviors of Students in Terms of Standardized Measures and Scales.

Standardized achievement and ability tests, the scores on which allow inferences to be made regarding the extent to which cognitive objectives concerned with knowledge, comprehension, understandings, skills, and applications have been attained.

Standardized self inventories designed to yield measures of adjustment, appreciations, attitudes, interests, and temperament from which inferences can be formulated concerning the possession of psychological traits (such as defensiveness, rigidity, aggressiveness, cooperativeness, hostility, and anxiety).

Standardized rating scales and check lists for judging the quality of products in visual arts, crafts, shop activities, penmanship, creative writing, exhibits for competitive events, cooking, typing, letter writing, fashion design, and other activities.

Standardized tests of psychomotor skills and physical fitness.

## II. Indicators of Status or Change in Cognitive and Affective Behaviors of Students by Informal or Semiformal Teacher-made Instruments or Devices.

Incomplete sentence technique: categorization of types of responses, enumeration of their frequencies, or ratings of their psychological appropriateness relative to specific criteria.

Interviews: frequencies and measurable levels of responses to formal and informal questions raised in a face-to-face interrogation.

Peer nominations: frequencies of selection or of assignment to leadership roles for which the sociogram technique may be particularly suitable.

Questionnaires: frequencies of responses to items in an objective format and numbers of responses to categorized dimensions developed from the content analysis of responses to open-ended questions.

Self-concept perceptions: measures of current status and indices of congruence between real self and ideal self -- often determined from use of the semantic differential or Q-sort techniques.

*Appended material to paper entitled "Paradigm Involving Multiple Criterion Measures for the Evaluation of the Effectiveness of School Programs" presented at the 1967 Annual Meeting of AERA, February 16, 1967, held in New York City.

Self-evaluation measures: student's own reports on his perceived or desired level of achievement, on his perceptions of his personal and social adjustment, and on his future academic and vocational plans.

Teacher-devised projective devices such as casting characters in the class play, role playing, and picture interpretation based on an informal scoring model that usually embodies the determination of frequencies of the occurrence of specific behaviors, or ratings of their intensity or quality.

Teacher-made achievement tests (objective and essay), the scores on which allow inferences regarding the extent to which specific instructional objectives have been attained.

Teacher-made rating scales and check lists for observation of classroom behaviors: performance levels of speech, music, and art; manifestation of creative endeavors, personal and social adjustment, physical well being.

Teacher-modified forms (preferably with consultant aid) of the semantic differential scale.

## III. Indicators of Status or Change in Student Behaviors Other than Those Measured by Tests, Inventories, and Observation Scales in Relation to the Task of Evaluating Objectives of School Programs

Absences: full-day, half-day, and other selective indices pertaining to frequency and duration of lack of attendance.

Anecdotal records: critical incidents noted including frequencies of behaviors judged to be highly undesirable or highly deserving of commendation.

Appointments: frequencies with which they are kept or broken.

Articles and stories: numbers and types published in school newspapers, magazines, journals, or proceedings of student organizations.

Assignments: numbers and types completed with some sort of quality rating or mark attached.

Attendance: frequency and duration when attendance is required or considered optional (as in club meetings, special events, or off-campus activities).

Autobiographical data: behaviors reported that could be classified and subsequently assigned judgmental values concerning their appropriateness relative to specific objectives concerned with human development.

Awards, citations, honors, and related indicators of distinctive or creative performance: frequency of occurrence or judgments of merit in terms of scaled values.

Books: numbers checked out of library, numbers renewed, numbers reported read when reading is required or when voluntary.

Case histories: critical incidents and other passages reflecting quantifiable categories of behavior.

Changes in program or in teacher as requested by student: frequency or occurrence.

Choices expressed or carried out: vocational, avocational, and educational (especially in relation to their judged appropriateness to known physical, intellectual, emotional, social, aesthetic, interest, and other factors).

Citations: commendatory in both formal and informal media of communication such as in the newspaper, television, school assembly, classroom, bulletin board, or elsewhere (see Awards).

"Contracts": frequency or duration of direct or indirect communications between persons observed and one or more significant others with specific reference to increase or decrease in frequency or to duration relative to selected time intervals.

Disciplinary actions taken: frequency and type.

Dropouts: numbers of students leaving school before completion of program of studies.

Elected positions: numbers and types held in class, student body, or out-of-school social groups.

Extracurricular activities: frequency or duration of participation in observable behaviors amenable to classification such as taking part in athletic events, charity drives, cultural activities, and numerous service-related avocational endeavors.

Grade placement: the success or lack of success in being promoted or retained; number of times accelerated or skipped.

Grade point average: including numbers of recommended units of course work in academic as well as in non-college preparatory programs.

Grouping: frequency and/or duration of moves from one instructional group to another within a given class grade.

Homework assignments: punctuality of completion, quantifiable judgments of quality such as class marks.

Leisure activities: numbers and types of; times spent in; awards and prizes received in participation.

Library card: possessed or not possessed; renewed or not renewed.

Load: numbers of units or courses carried by students.

Peer group participation: frequency and duration of activity in what are judged to be socially acceptable and socially undesirable behaviors.

Performance: awards, citations received; extra-credit assignments and associated points earned; numbers of books or other learning materials taken out of the library; products exhibited at competitive events.

Recommendations: numbers of and judged levels of favorableness.

Recidivism by students: incidents (presence or absence or frequency of occurrence) of a given student's returning to a probationary status, to a detention facility, or to observable behavior patterns judged to be socially undesirable (intoxicated state, dope addiction, hostile acts including arrests, sexual deviation).

Referrals: by teacher to counselor, psychologist, or administrator for disciplinary action, for special aid in overcoming learning difficulties, for behavior disorders, for health defects or for part-time employment activities.

Referrals: by student himself (presence, absence, or frequency).

Service points: numbers earned.

Skills: demonstration of new or increased competencies such as those found in physical education, crafts, homemaking, and the arts that are not measured in a highly valid fashion by available tests and scales.

Social mobility: numbers of times student has moved from one neighborhood to another and/or frequency which parents have changed jobs.

Tape recordings: critical incidents contained and other analyzable events amenable to classification and enumeration.

Tardiness: frequency of.

Transiency: incidents of.

Transfers: numbers of students entering school from another school (horizontal move).

Withdrawal: numbers of students withdrawing from school or from a special program (see Dropouts).

IV. Indicators of Status or Change in Cognitive and Affective Behaviors of Teachers and Other School Personnel in Relation to the Evaluation of School Programs.

Articles: frequency and types of articles and written documents prepared by teachers for publication or distribution.

Attendance: frequency of, at professional meetings or at in-service training programs, institutes, summer schools, colleges and universities (for advanced training) from which inferences can be drawn regarding the professional person's desire to improve his competence.

Elective offices: numbers and types of appointments held in professional and social organizations.

Grade point average: earned in postgraduate courses.

Load carried by teacher: teacher-pupil or counselor-pupil ratio.

Mail: frequency of positive and negative statements in written correspondence about teachers, counselors, administrators, and other personnel.

Memberships including elective positions held in professional and community organizations: frequency and duration of association.

Model congruence index: determination of how well the actions of professional personnel in a program approximate certain operationally-stated judgmental criteria concerning the qualities of a meritorious program.

Moonlighting: frequency of outside jobs and time spent in these activities by teachers or other school personnel.

Nominations by peers, students, administrators, or parents for outstanding service and/or professional competencies: frequency of.

Rating scales and check lists (e.g., graphic rating scales or the semantic differential) of operationally-stated dimensions of teachers' behaviors in the classroom or of administrators' behaviors in the school setting from which observers may formulate inferences regarding changes of behavior that reflect what are judged to be desirable gains in professional competence, skills, attitudes, adjustment, interests, and work efficiency; the perceptions of various members of the total school community (parents, teachers, administrators, counselors, students, and classified employees) of the behaviors of other members may also be obtained and compared.

Records and reporting procedures practiced by administrators, counselors and teachers: judgments of adequacy by outside consultants.

Termination: frequency of voluntary or involuntary resignation or dismissals of school personnal.

Transfers: frequency of requests of teachers to move from one school to another.

V. Indicators of Community Behaviors in Relation to the Evaluation of School Programs

Alumni participation: numbers of visitations, extent of involvement in PTA activities, amount of support of a tangible (financial) or a service nature to a continuing school program or activity.

Attendance at special school events, at meetings of the board of education, or at other group activities by partens: frequency of.

Conferences of parent-teacher, parent-counselor, parent-administrator sought by parents: frequency of request.

Conferences of the same type sought and initiated by school personnel: frequency of requests and record of appointments kept by parents.

Interview responses amenable to classification and quantification.

Letters (mail): frequency of requests for information, materials, and servicing.

Letters: frequency of praiseworthy or critical comments about school programs and services and about the personnel participating in them.

Participant analysis of alumni: determination of locale of graduates, occupation, affiliation with particular institutions, or outside agencies.

Parental response to letters and report cards upon written or oral request of school personnel: frequency of compliance by parents.

Telephone calls from parents, alumni, and from personnel in communications media (e.g., newspaper reporters): frequency, duration, and quantifiable judgments about statements monitored from telephone conversations.

Transportation requests: frequency of.

# FORMAT FOR AN EVALUATION REPORT FOR AN EDUCATIONAL PROGRAM

## SECTION I -- OBJECTIVES OF THE EVALUATION

A. Audiences to be Served by the Evaluation

B. Decisions about the Program, Anticipated

## SECTION II-- SPECIFICATIONS OF THE PROGRAM

A. Educational Philosophy behind the Program

B. Subject Matter

C. Learning Objectives, Staff Aims

D. Instructional Procedures, Tactics, Media

E. Students

F. Instructional and Community Setting

G. Standards, Bases for Judging Quality

## SECTION III--PROGRAM OUTCOMES

A. Opportunities, Experiences Provided

B. Student Gains and Losses

C. Side Effects and Bonuses

D. Costs

## SECTION IV-- RELATIONSHIPS AND INDICATORS

A. Congruence

B. Contingencies

C. Trend Lines, Indicators, Ratios

## SECTION V -- JUDGMENTS OF WORTH

A. Value of Outcomes

B. Relevance of Objectives to Needs

C. Usefulness of Evaluation Information Gathered

Stake
7/29/69

# EVALUATION REPORTS

## DESCRIBING THE CONTEXT OF A PROGRAM

### City or Community Characteristics

What is the population of the city or community?
What adjective(s) would typically be used to describe the city or community?
In what part of the country is it located?
What is the percentage of deteriorating or dilapidated housing in the city or community?
What is the city- or community-wide unemployment rate?
What percent of families in the city or community are on welfare?
What is the city- or community-wide literacy rate?
What is the city- or community-wide school dropout rate?
What is the city- or community-wide delinquency rate?
Are there any special educational problems faced by the city or community?
What attempts, if any, are being made to deal with these problems?

### Neighborhood characteristics

What adjective(s) would typically be used to describe the neighborhood(s)?
What is the average family income in the neighboorhood(s)?
What is the literacy rate in the neighborhood(s)?
What kinds of occupations do most of the people in the neighborhood(s) have?
What is the unemployment rate of the neighborhood(s)?
What percent of the families in the neighborhood(s) are on welfare?
What is the percent of nonintact families in the neighborhood(s)?
What ethnic groups, in what percent, are represented in the neighborhood(s)?
What linguistic groups, in what percent, are represented in the neighborhood(s)?
What is the population density (number of people per square mile) in the in the neighborhood(s)?
What is the percent of multi-family dwellings in the neighborhood(s)?
What percent of the dwellings were built pre-1940 in the neighborhood(s)?
What percent of the dwellings are rental (rather than owner-occupied) in the neighborhood(s)?
What is the percent of deteriorating or dilapidated housing in the neighborhood(s)?
What is the school dropout rate in the neighborhood(s)?
What is the delinquency rate in the neighborhood(s)?
Have these neighborhood characteristics remained constant in the last few years or is the neighborhood(s) in transition?

### School Characteristics - General

What was the per capita expenditure, including both capital and operating expenses, prior to the program?
What was the salary range for teachers in the school(s) for the year immediately preceding the program?
What is the age and condition of the main school building(s)?
What grade levels were included in the school(s)?
What was the average teacher-pupil ratio in the school(s)?
How were the students routinely grouped in the school(s)?
Were any pupils enrolled in the school(s) as a result of a bussing or open enrollment program?

Prepared by C. O. Neidt, Colorado State University, 1969.

## School Characteristics - General (continued)

Was a conventional curriculum followed in the school(s)?
What services, personnel, or special programs were available in the school(s) prior to the program?
Were any other specially funded programs ongoing in the school(s) prior to the beginning of this program?
At what intervals are achievement tests routinely given?
What achievement tests are routinely given? To what grades?
How are these achievement tests administered and by whom?
How did the achievement level of the school(s) compare with city-wide and/or national norms prior to the program?

## School Characteristics - Teachers

What were the paper qualifications of the teachers?
What was the average number of years of teaching experience?
What was the average age of the teachers?
What was the male-female ratio of teachers?
What ethnic groups, in what percent, were represented by the teachers?
What linguistic groups, in what percent, were represented by the teachers?
What was the teacher turnover in the school(s) prior to the beginning of the program?

## School Characteristics - Student Body

What was the pupil enrollment in the school(s) at the beginning of the academic year?
How many pupils withdrew or transferred from the school(s) after the school year began?
How many pupils enrolled in the school(s) after the school year began?
What was the average daily attendance in the school(s)?
Has the total pupil enrollment in the school(s) involved inthe program changed in the last three years?
What ethnic groups, in what percent, were represented by the students?
What linguistic groups, in what percent, were represented by the students?
What was the male-female ratio of the students?

## Historical Background

Did the program exist prior to the time period covered in the present report?
Is the program a modification of a previously existing program?
How did the program originate?
What special efforts were made to gain acceptance of the pgroram by parents and the community before it began?
If special problems were encountered in gaining acceptance of the program by parents and the community, how were these solved so that the program could be introduced?

# DESCRIBING THE TREATMENT PROVIDED BY A PROGRAM

## Personnel: Instructional and Noninstructional

What categories of personnel were added by program?
What regular staff were assigned to program?
What new staff were hired for program?
What were paper qualifications for various personnel?
What were average years of relevant experience of personnel?
What were the most important duties of personnel?
What was the time commitment of various personnel?
What in-service training was provided?
What was the male-female ratio of classroom personnel?
What personnel characteristics enhanced or reduced program effectiveness?
How did special needs of pupils affect staff development and utilization?

## Supporting Services

What services were part of the program?
What services were available to experimentals? To controls? To both?
How did special needs of pupils affect provision of services?

## Organization: Schedules

For how long did the program operate?
How were experimental and control classes scheduled in the total school context?
How many hours of instruction did experimentals receive? Controls?
Were time intervals between learning and testing equivalent for these groups?

## Organization: Planning

Were meetings held regularly for experimental and control teachers?
What were the purposes of these meetings?
Who was present (besides teachers) and why?

## Organization: Physical Arrangements

Where were experimental classes located?
Where were control classes located?
What were the most noteworthy features of physical arrangements in each?

## Organization: Grouping of Teachers

How were experimental and control teachers grouped for instructional purposes?

## Organization: Grouping of Pupils

How were pupils grouped within the total school context?
How were pupils grouped for instruction in experimental and control classes?
How many children were in each experimental class? In each control class?

## Major Program Segments

What major segments comprised program?
Which of these were available to experimentals? To controls? To both?
Were segments equivalent for these groups in the following respects:
  Objectives?
  Emphasis?
  Provision for motivating pupils?
How did special needs of pupils affect content of major program segments?
What characteristics of these segments enhanced or reduced program effectiveness?

## Methodology: Pupil Activities

What were main activities of experimentals? Of controls?
How much time was devoted to each main activity?
How many pupils were involved in each?
How were instructional materials used by pupils in each?
Did pupils have freedom of choice in participating in each main activity?
How much time did pupils spend in the program each day? Each week?

## Methodology: Teacher Activities

What were main activities of teachers in experimental and control classes?
How much time did the teacher spend with the pupils?
What was the teacher-pupil ratio (or aide- or adult-pupil ratio)?
What provision did the teacher make for pupil response?
How did the teacher use various instructional materials for the activity?
What provision did the teacher make for individualizing instruction?
To what extent were teachers free to experiment with teaching methods?
How did the teacher give feedback to pupils on individual progress?
What provision did the teacher make for motivating pupils?
Were amounts of practice, review, and quiz activities equivalent for both groups?
Was content of these activities equivalent for both groups?
How did special needs of pupils affect teaching methods?
What characteristics of activities enhanced program success?

## Instructional Equipment and Materials

What equipment and materials were used by experiemntals? Controls? Both?
In what amounts?
What equipment and materials were used in each main activity in the two groups?
What specific features suited a given device to a particular activity?
Were materials equivalent for both groups in the following respects:
  Subject-matter content?
  Content of drill?
  Vocabulary level?
What instructional materials were developed for program? How were they developed?
What characteristics of materials enhanced or reduced program effectiveness?
How did special needs of pupils affect selection or development of materials?

## Parent-Community Involvement

What provisions were made for parent and/or community involvement in the program?
Were these provisions equivalent for parents of experimentals and controls?
Were group meetings and/or parent conferences held for parents of experimentals and controls? Describe.

## Budget

What was the total cost of program? (indicate length of time covered)
From what sources were these funds obtained?
What portion of total program cost was start-up expense? Continuation expense?
Can you break down total program cost into broad categories of expenses?
If the program were repeated, how would you modify the budget?
What was per-pupil cost of program?
How does it compare with normal per-pupil cost of schools in the program?
Where can the reader get additional budget information?

## DESCRIBING, ANALYZING AND INTERPRETING EVIDENCE OF CHANGES INDUCED BY A PROGRAM

### Objectives:

What was the program aiming to do for the children and adults in it?

Were the children expected to improve their scores on achievement measures? If so, in what areas?
Were the teachers or other adults expected to change their modes of instruction?
Were the children expected to change their attitudes? If so, which ones?
Were the teachers or other adults expected to change their attitudes? If so, which ones?

### Sampling Procedures:

How were the children and adults in the program chosen?

Were the samples originally representative of the populations from which they were chosen?
Were the controls selected before or after the program?
Were steps taken to avoid the samples being affected by other programs?
Were steps taken to avoid real differences in the quality of teachers selected for experimental and control groups?
Was there attrition of the samples?
Was there attrition of groups of children with the same characteristics?
Were pupils added to the samples to replace dropouts?
Were there many children who did not receive the treatment often because of poor attendance?
Did the children participate voluntarily?
Were the same children included in both pretest and posttest samples?

Describing Samples:

Which children received the treatment, from which adults?

What is the size of the experimental sample?
What is the age or grade level of the experimental sample?
How is the experimental sample divided into boys and girls?
Are achievement scores available by which to describe the experimental sample?
Which adults gave the treatment that constituted the program?

Measuring Change:

What measures were applied to find out whether the program's aims had been achieved?

Were the measures matched to the objectives in content?
Did the tests used have sufficient "floor" and "ceiling"?
Were the same measures used for both experimental and control groups?
Were the same measures (or parallel forms) used for both pre and posttesting?
Were IQ tests used when achievement tests were more appropriate?
Was the reliability of the tests quoted?
Under what conditions were the measures applied?
Were the same or different testers used for successive testings?
Were oral, or written, instructions available for the tests?
Were assessors or observers likely to bias the results for or against the program?
How much time elapsed between testings?
Were assessors or observers specially trained?

Presenting Data:

What data were obtained from the measures applied?

What measures of central tendency should be used?
What measures of dispersion were used?
Were there graphical displays which could have been used to present data more clearly?

Analyzing Data:

What analyses were undertaken of the data?

Was there a proper basis against which to compare the progress of the experimental group?
What was the correlation between pretest and posttest?
What comparisons were drawn for subsamples?
Is there any evidence that children who attended more gained more from the program?
Was the formula or source given for the statistical test applied?
Did the data meet the prerequisites for the statistical tests used?
Were there real differences between the groups?

Drawing Conclusions:

What conclusions were drawn from the analyses of the results?

Were the conclusions based on statistical probability?
Were the statistical conclusions translated into ordinary language?
Were other conclusions stated in ordinary language?
Can the conclusions be generalized, or are they applicable only to the sample or population served by the program?
Were the conclusions of educational importance?
What recommendations can be based upon the conclusions?

## AERA MONOGRAPH SERIES ON CURRICULUM EVALUATION

### *Now Available*

No. 1 ($2)* *Perspectives of Curriculum Evaluation*

Ralph Tyler, "Changing Concepts of Educational Evaluation"
Robert Gagné, "Curriculum Research and the Promotion of Learning"
Michael Scriven, "The Methodology of Evaluation"

No 2 ($2)* *Evaluation Activities of Curriculum Projects*
by Hulda Grobman

No. 3 *Instructional Objectives*
Elliot Eisner, "Instructional and Expressive Educational Objectives: Their Formulation and Use in Curriculum"
W. James Popham, "Objectives and Instruction"
Howard Sullivan, "Objectives, Evaluation, and Improved Learner Achievement"
Louise Tyler, "A Case History: Formulation of Objectives from a Psychoanalytic Framework"

### *In Press*

No. 4 *Research Strategies for Evaluating Training*
Papers edited by Philip DuBois and Douglas Mayo

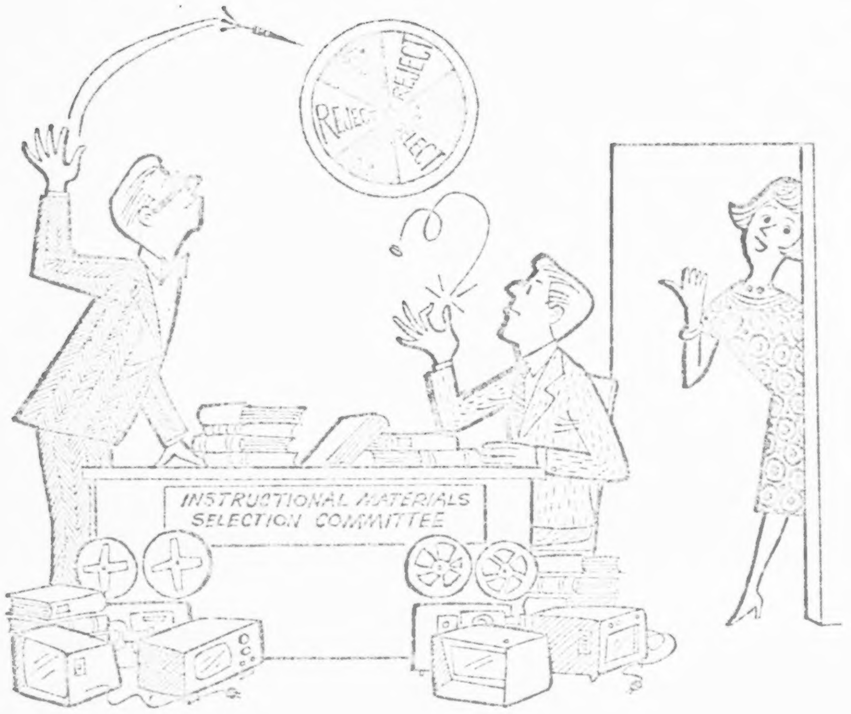
No. 5 *Evaluation as a Tool in Curriculum Development: The IPI Evaluation Program*
by C. M. Lindvall and Richard Cox

No. 6 *Classroom Observation*
Graham Nuthall, "A Review of Some Selected Recent Studies of Classroom Interaction and Teaching Behavior"
James Gallagher, "A Topic Classification System for Classroom Observation"
James Gallagher, "Three Studies of the Classroom"
Barak Rosenshine, "Some Criteria for Evaluating Category Systems: An Application to the Topic Classification System"

Available from: Rand McNally and Company
P. O. Box 7600
Chicago, Illinois 60680

*Price for soft-bound copy. Hard-bound copy and AERA-membership prices available.

PURCHASING
ISN'T
FUN AND
GAMES



EPIE HELPS
SCHOOLS
PURCHASE
WISELY

“ *Educational Product Report.* Monthly, October-June. $35. Educational Products Information Exchange Institute, 386 Park Avenue South, New York, N. Y. 10016.

Published by the EPIE Institute, a non-profit, nonindustry, nongovernment professionals' cooperative which conducts impartial studies of the effectiveness of instructional materials, equipment, and systems. This unique publication, EPIE's major outlet for the results of its nationwide cooperative research on educational products, contains comprehensive data and articles on a different product area each month. It impresses one as a much-needed consumer's report for the educational decision-maker. Its value lies not only in the fact that it is the only educational publication of its kind, but in its careful research, clear editing, handsome appearance, and, above all, in its usefulness. Recent issues have dealt with programmed instruction, science kits, 8mm projectors, and study carrels, and others have covered mathematics, social studies and science textbooks, classroom television receivers, and standards for audio-visual equipment. "Hard" comparative data as to price, grade level, content or design, producer's warrantees, etc. are presented in a self-contained section, and informative articles, interviews, analyses, and guidelines are presented in the main section of each issue. While reports on materials are limited to the elementary and high school grades, those on equipment are clearly useful to colleges. Some (such as the report on study carrels) are specifically useful to libraries. The EPIE Institute, its research program, and *Educational Product Report* are supported almost solely by its subscribing members (no advertising or funding from producers is accepted). A subscribing membership costs $35 and is a solid investment in good decision-making. (Note: Our February 15th column reviewed *Educational Product Report* by its previous title, *The EPIE Forum,* and erroneously reported that it was free. We particularly regret this error, as this nonprofit cooperative is obviously dependent on the $35 subscribing membership fee it receives from the schools and libraries it serves.) ”

*Reprinted with permission. Library Journal, May 15, 1969*

TO BECOME A SUBSCRIBING-MEMBER OF EPIE's EDUCATIONAL PRODUCT REPORT, PLEASE COMPLETE AND MAIL THE ATTACHED (POSTAGE-PAID) SUBSCRIPTION-MEMBERSHIP FORM.

EPIE **The Educational Products Information Exchange Institute**
386 Park Avenue South, New York, N. Y. 10016

Please enter one (1) REGULAR Subscribing-Membership @ $35.00 for the '69-'70 school year (monthly October-June).

Your Name and Title ______________________

School ______________________

Street ______________________

City ______________ State ________ Zip Code ________

☐ Please make this a RETROACTIVE SUBSCRIBING-MEMBERSHIP for which we will receive all ('67-'68, '68-'69) REPORTS and agree to pay an additional $35.00.

☐ If separate purchase order to follow, please check here.

### SPECIAL RETROACTIVE SUBSCRIBING-MEMBERSHIP

At your option, you may receive all ('67-'68, '68-'69) issues of EPIE's Educational Product Report from its first (September, 1967) through June, 1969.

Comprehensive Reports on:

- Overhead Projectors
- Modern Math
- Science Kits
- Science Texts
- T.V. Receivers
- 8mm Projectors
- Social Studies
- Black History & Biography
- Study Carrels
- Evaluation Tools
- AV Standards
- Programmed Instr'n
- Phonics Books

# APPENDIX E: AERC Evaluation Workshop Case Studies Nos. 1-6

## AERC EVALUATION WORKSHOP

### Case Study No. 1

### Evaluation for Evening Class Personnel Utilization

Franklin Community College is located in Garnet City, the county seat of Grassland County, situated in the center of an east-Central state. Garnet City is 150 miles from the nearest large city. Its population is 240,000, while the population of the county is 410,000, and the trade area is 755,000.

This area is serviced by two state universities, one state teachers college, one private university, two private colleges, Franklin Community College, and a variety of small propriety and trade schools.

The Continuing Education Division of Franklin Community College was established twelve years ago, primarily teaching evening courses for the first five years. Approximately 5,400 students are presently serviced by the division. It is organized by instructional format into three bureaus: Class, conference, and correspondence. There is a full-time director of the division who reports to the President of the college.

Ernest Trueblood, with an M.A. in Business Teacher Education, is the half-time program administrator for 50 sections of business courses in the Continuing Education Division (CED) of Franklin Community College. During half of his time, he teaches in the day program of the college's Department of Business Administration, and for the other half of his time he has been developing an increasingly large and effective evening class program of business courses in the CED. His program reaches 1500 different adults, 1100 in credit courses and 400 in non-credit courses. In the past year, his staff numbered 62 teachers. Half of these were from the community. His current decisional problem is the appointment of Justinian Eagle, a local attorney with no previous connection with Franklin Community College, to teach an advanced evening credit course next fall on Business Law, with an emphasis on the uniform commercial code. Mr. Eagle is a Vice-President of the Garnet City National Bank, and about fifteen years ago had taught a course in business law at the Law School of the University in the state capital, 150 miles away.

The teacher of the business law course in the day program for the past few years has been a young faculty member by the name of Casper Wilton. Business Law is not a major area of competence or interest of Mr. Wilton, but there was no one else in the Department of Business Administration who was more prepared and interested, so he agreed to teach it as a favor to Dr. Harry Slick, the Chairman of the Department of Business Administration in the daytime program of the college. Dr. Slick recruited Mr. Wilton to Franklin, in part because of his extensive publications for a young man and his growing scholarly reputation. Mr. Wilton is currently working on his doctorate in Economics, and he is interested in teaching the business law course next fall on an overload basis for extra pay.

In the classroom, Mr. Wilton tends to be hesitant and somewhat disorganized. He has had little practical business experience. His one teaching experience in the Bureau of Evening Classes several years ago was not very successful, and there was a 60 percent drop out rate for the adult students in the course. When Ernest Trueblood first approached Dr. Slick about offering the Business Law course through the Bureau and having Mr. Eagle teach it, Dr. Slick countered that it was a good idea to offer the course, but that if it was offered, Casper Wilton should teach the course. In part because he reports to Dr. Slick in the teaching half of his assignment, Mr. Trueblood hesitated to take issue with Dr. Slick's seemingly firm position at their first encounter on the topic.

Mr. Trueblood subsequently visited informally about the matter with Mr. Arlin Marlin, the Director of the CED and with Dr. August Steele, the Director of the Business and Commercial Division of the preparatory education program for full-time students at the college. Ernest stressed to Mr. Marlin the backlog of interest and requests from the community for a business law course, and the fact that many of the potential class members are businessmen of some influence in Garnet City, especially in relation to authorizing tuition reimbursement arrangements for employees who enroll in the CED. Mr. Marlin indicated that in the past, the decision on staffing credit courses has been a mutual one and that there is little that can be done if Dr. Slick insists on Wilton instead of Eagle, other than not offering the course, at least on a credit basis. Mr. Marlin, however, did indicate that he was sympathetic with Mr. Trueblood's views on the matter.

Ernest stressed to Dr. Steele the importance of the close working relationships between his division of the college and the local business community in terms of preparatory education, placement of graduates, continuing education, research, consultation, and the inclusion of qualified practitioners in teaching roles. Dr. Steele indicated that the decision was up to Dr. Slick and that he would support whatever decision Dr. Slick made.

Dr. Steele acknowledged the validity of both arguments and after conferring with Mr. Marlin urged Ernest Trueblood and Mr. Slick to come to you for assistance in providing information relevant to the decision that Mr. Marlin and Mr. Steele must make concerning who is to teach the course.

# AERC EVALUATION WORKSHOP
## Case Study No. 2
## Evaluation of Adult Basic Education Materials

More than 100,000 of the million people who live in Central City are adults whose educational achievement is less than the equivalent of eighth grade. Many of the adults are functionally illiterate in that they are unable to read and write at a minimum level needed for economic and social functioning. Rough estimates are that the target population consists of 30,000 native born blacks most of whom are under 40; about 25,000 immigrants, most of whom are between 35 and 70; about 20,000 native born whites most of whom are between 25 and 60; about 10,000 Spanish speakers most of whom are under 30; and the remainder is even more varied. There is a wide range of age, ability, and economic level. It is commonly believed that many of them would be better able to function in society if they had a higher level of literacy, at least equal to the equivalent of eighth grade.

About five years ago, ABE classes were begun which were paid for primarily by federal funds authorized under Title III of the Elementary and Secondary Education Act. As the number of ABE classes increased, an ABE office was organized in the adult education division of the schools. The original ABE coordinator, a former elementary school assistant principal by the name of Wiley Wilson, is the current coordinator of the ABE office. Mr. Wilson's annual report last year listed more than 700 ABE participants in about 50 ABE classes held in evening schools for an average enrollment of more than 15. Each school has a teacher-in-charge who coordinates the program. In addition about 300 ABE participants attend 30 classes held in non-school facilities. Most of the ABE teachers are certified teachers of the Central City schools who in addition teach one ABE course each term that meets one evening a week. Wiley Wilson is the only person who works full-time on ABE.

Wiley Wilson's problem is a policy issue regarding materials development and selection. In past years materials selection has been left almost entirely to the individual teacher. As a teacher decided on commercial materials to be purchased, she placed an order with Mr. Wilson's office and he tried to obtain them. Most of the teachers have limited experience working with adults. It has become clear to Wiley that most of the elementary education materials for the teaching of reading and arithmetic have limited usefulness with adults. Few of the ABE teachers have attempted to develop their own materials for adults and those materials which have been developed have been of poor quality. There has been a great increase during the past few years in the amount of commercial ABE materials on the market, but there is a great variability in their quality. Wiley Wilson and the others connected with the ABE program do not have enough background in the area of materials selection to know exactly what to choose, but they are convinced that the present arrangement is unsatisfactory and that something needs to be done.

Many of the adult learners have complained about the lack of relevance and adult orientation of the present materials. It is not known what basis individual teachers are using for the selection of materials. With the diffuseness of ABE goals, it appears that decisions regarding materials selection are having a major influence on the shaping of the instructional program in each class. Although there is a wide range of materials within the total ABE program, the limited range within each class makes individualization of instruction difficult. For the ABE teacher who wants to develop her own materials, there is little by way of facilities, equipment, and staff assistance to help her do so effectively.

The problem is illustrated by the typical practices of two of the ABE teachers with whom Wiley has talked about the problem, Miss Plotz and Miss Zing. Miss Plotz uses mostly commercially published materials. She selects from those with which she is familiar as an elementary school teacher and those which seem to be most appropriate for use with adults. Several terms ago, she sent a request to Wiley to order a substantial amount of consumable materials such as workbooks for each of the participants in her class. Wiley replied that the materials budget was not adequate for that purpose so she ordered a few more books and has not tried to reorder consumable materials.

Miss Zing prepares most of her own materials. Because the ABE program lacks both duplicating equipment of its own and convenient arrangements for having materials duplicated, Miss Zing obtains ditto masters from her elementary school and runs copies on the machine there for use with her ABE class. She has adapted materials that she had prepared earlier for use with her children, and also commercial materials. She also exchanged materials with a few other ABE teachers. Miss Zing expressed the opinion that although her teacher-made materials are probably better than most of the child-oriented materials with which she is familiar, that with expert assistance they could probably be improved substantially.

The major task that confronts Wiley Wilson is how can he best proceed to improve the process of ABE materials selection? His tentative conclusion is that the ABE program needs a general procedure for evaluating existing and proposed materials. If you were confronted with the task of developing a plan for evaluation of ABE materials, what would your plan look like.

# AERC EVALUATION WORKSHOP
## Case Study No. 3
## Evaluation of an Off-Campus Extension Program

State University is the largest institution of higher learning and the major one supported by public funds in one of the southwestern states. Its main campus is located in Collegeville, twelve miles from the capital and major city. In 1957 the Extension Course Department of the Extension Division was established to aid and plan non-credit courses throughout the state. Although several buildings form a campus in Maintown, the state's second largest city 300 miles from Collegeville, most of the Extension Division's non-credit classes meet wherever they can find space in small towns in this largely rural state. The Extension Division's staff is made up of ten supervisors, two of whom travel around the state and supervise 150 part-time teachers of non-credit classes. (This does not include the faculty at Maintown.)

Hickston is a town in the center of the state, 120 miles from Collegeville. Nine thousand people live in the town and another 5,000 live outside of it, but use it as their main shopping area. It is a county seat and has two elementary schools, a junior high school, and a senior high school. Of the 14,000 inhabitants, 75% are white, with the remaining 25% made up largely of blacks, Mexican immigrants, and American Indians. The average family income is $5,500 a year, with few families considered to be on a poverty level. Most of the people in town are employed in a large grain elevator, a broom factory, the local schools, or own or work in the local shops. Most who live on the outskirts own small farms and ranches.

State University's Extension Division, Department of Extension Courses, has offered the following courses (all non-credit) in Hickston: decorating indoors with plans, creative writing workshop, decorative candle making, creative forms of stichery, principles of photography, small business management, English improvement for the foreign born. The number of courses varies from year to year depending on how many teachers are available. All teachers have been local 'experts' supervised by Chester Chaff, one of the two traveling staff from Collegeville.

This year Chester Chaff has found the enrollment in the five courses offered at a new low. In the past from 250-500 people have taken part in some phase of the program each semester. Only 140 people are registered at present, and the attendance rate drops each week. Below is a breakdown of the courses and the attendance figures:

| Class | No. Regist. | Ses. 1 | Ses. 2 | Ses. 3 |
|---|---|---|---|---|
| Writing Workshop | 20 | 19 | 15 | 16 |
| Photography | 32 | 32 | 26 | 25 |
| Business Management | 29 | 25 | 22 | 23 |
| Knitting | 40 | 30 | 31 | 27 |
| English for the Foreign Born | 17 | 16 | 15 | 16 |
|  | 138 | 122 | 109 | 107 |

The teachers who used to be paid about $500 - $750 a semester demanded a salary increase and are now being paid $750 - $900 a term. Therefore, the tuition has increased from $25 to $40 for a twelve-week course. Materials like cameras, film, and yarn are additional expenses for the students who use them. Another problem is the classes which used to meet in a modern elementary school now meet in the older, poorly equipped junior high school. Chaff had to make the move because he could no longer obtain janitorial services at the elementary school. The new site has poor parking facilities as well as a less pleasant interior. Try as he might, Chester has been unable to arrange for facilities for his classes, at as low a rate as he pays for the use of the junior high school.

A group of young adults who live and work in the community have urged Chaff to offer some credit courses which they will be able to use if they decide to attend the University in the future. Chaff has wanted to do this, but he has been unable to find any qualified local personnel who are willing to teach such courses, and few faculty from the main campus have been willing to travel the 120 miles for the present payment arrangement. A fourth problem which has come to Chaff's attention is that one of his instructors, although familiar with the area in which he is teaching, is a poor classroom instructor. He is not liked by his students, and many have dropped out. One other reason for the generally high drop-out rate is that a flu epidemic has hit the town, and many of the women have had to remain at home to nurse their children.

Although the last problem might remedy itself in future semesters, Chaff is not optimistic about the over-all picture for non-credit Extension Education in Hickston. He does not like the thought of having to discontinue classes there, but unless he makes some changes which become obvious to the community, Chaff fears he will have to end the operation in this town. Chaff has come to you to ask your assistance in designing a plan to gather information relevant to the decision with which he is faced.

# AERC EVALUATION WORKSHOP
## Case Study No. 4
## Evaluation of Institutional Change

Walkersville is a small city in New England. Several hundred thousand pieces of mail pass through its post office every month. Central Post Office is in the center of town, and there are two other auxiliary post offices, one in the city's north and one in the south. Between seventy-five and one hundred people are employed at Central Post Office, including the letter carriers.

Wallace Handstamp, the Post Master of the town, rules the post office with an iron hand. He is quick to anger and quick to make decisions. Handstamp has been a civil servant working for the post office for over twenty years. In the past three years since he has been Post Master, he has made few improvements, not willing to give up the ways he has grown accustomed to. Recently Handstamp decided to add a new automatic sorter to the other machinery in the post office. He only agreed to do this after much pressure from Pierre Zip, the head sorter.

Zip, who recently became an American citizen, used to work in the Bureau de Poste in his native France. He is anxious to add the most modern machines and adapt the newest ideas in the post office. Zip promised Handstamp that he would train the men who would operate the new sorter himself, and he has made several trips to the state capital to learn how the machine works. Handstamp, although agreeing to order the machine, told Zip that he did not think that it would increase production enough to make up for the disruption of the habits of personnel doing related tasks.

Zip realizes that he must please Handstamp in order to keep his supervisory position. Although he has passed the required Civil Service exams, Handstamp could always move him to one of the smaller branch post offices. Zip recommended four men to be trained to operate the machine; the machine requires only one experienced man at a time, and these four could then train others.

Two older men, Joe Age and Fred Tired, and two younger men, Matt Young and Herb Youth, were selected to undergo the training program. They agreed because they felt pleased to be selected from the other workers and because the job, once learned, was relatively simple and would keep them off their feet. All of them were somewhat apprehensive about being trained by Zip. Small and slightly hunchbacked, Zip tended to be officious, too directive, and sometimes hard to understand. Although these four men had never had any difficulty in working with him, they knew others who had, and they looked to the new experience with mixed feelings.

Zip and Handstamp discussed the timing of the training sessions and discovered that they had very different ideas about the amount of time that was required. Zip had been thinking about eight half-days over a two week period, that would enable the four trainees to learn to operate the machine and also spend some time on human relations problems that would probably occur as the new machine affected related jobs, and on coaching techniques as they trained others to use the new machine. As Zip started to describe his plans to Handstamp, he discovered that Handstamp was thinking about four one-hour sessions scheduled at the end of four consecutive work days, partly on work-time and partly on the men's own-time.

At the end of a guarded conversation, a compromise was reached. The training sessions for the four men would consist of five, two-hour periods, all on company time, and would occur during a two-week period. Zip and the four men could decide on the specific days and times, and sometime could be devoted to matters other than the technical operation of the new machine.

Zip asked Handstamp if he would participate in part of the first session, and he agreed. The first training session began with a few comments by Zip and then an inspection of the new machine. Zip demonstrated its operation and added a few comments to provide a general overview of what was to be learned technically. He and the four men then went to an office where they could raise some questions that occurred to them. Then Handstamp joined them for the purpose of talking about how the new machine fit into the total operation of the post office. Zip was surprised and disappointed at the extent to which negative attitudes were expressed by Handstamp. He seemed to stress avoiding the disadvantages of the new machine and hardly mentioned benefiting from the advantages. After the first session, several of the men mentioned to Zip that the scuttlebut around Central Post Office was that Handstamp really didn't want the new machine and that if it didn't work out well, Zip would take the blame. The task facing Zip is to document the merits of the change that he has instigated. What plan for evaluating this change would you propose?

# AERC EVALUATION WORKSHOP

## Case Study No. 5

## Evaluation of Previous Cooperative Arrangements in Agricultural Extension

The agricultural extension agent in sparsely populated Sunflower County is Seaman Knapp. Yes, he's been ribbed about that since he enrolled as an extension education major at the College of Agriculture a decade ago. One of the changes that has occurred since the time of the Seaman Knapp was the shift from the county agent as an agricultural specialist, to agricultural generalist, to a current emphasis on being an educational generalist. This was illustrated by a phone call that he received just last week.

Knapp was in his office in the basement of the County Courthouse when he received a call from Cornelius Sodbuster, who a year earlier had moved into Sunflower County and had purchased a moderately small farm. Mr. Knapp knew that the farm had changed hands but had never met the new owner. "Corny", as he was known to his friends, had previously owned two farms elsewhere in the state, but had lost both due primarily to poor financial management. After he lost the second one, he rented a farm and also worked part-time in a nearby town to save money to buy his present farm. "Corny" expressed to Mr. Knapp his great determination to succeed.

Several days later Mr. Knapp visited "Corny" at his farm, as they had arranged during the initial phone call. "Corny" had called Mr. Knapp because he was weighing a decision about the purchase of minimum tillage equipment. He had heard such equipment allowed him to combine several steps involved in preparing the soil and planting. However, it was expensive and he was uncertain whether his farm was large enough or if it would be better to try to go in with several other nearby farmers to share the costs and the equipment.

As the two men looked over the farm and talked about the problem, it seemed to Mr. Knapp that the initial problem was only part of a larger problem of a viable plan that included capitalization and credit, anticipated costs of labor and supplies, anticipated income, and provision for reserves. "Corny" mentioned that earlier in the year he had left his wife and three teenage children at home and had enrolled in an evening agriculture course taught by the vocational agriculture teacher at the high school, but had dropped out because it didn't seem to be very relevant to his situation. He had recently come across a bulletin from the Experiment Station, but only a few parts were useful.

As Mr. Knapp left "Corny" he told him that they would talk together in a few days about the next steps. As Mr. Knapp drove back to his office, he thought about the best way that he could help "Corny". His own specialized background was in production aspects of farming such as agronomy, and his background in farm management was quite general. By the time that he arrived at his office, three other people came to mind who might singely or in combination be able to help "Corny".

One was "Curley" A. Dopter, a fairly successful young farmer who had been active in the local farmer's organization and who had assisted other farmers in working through this type of problem. Another was H. F. Seay, the county agent in the adjacent county whose specialized background included finance and agricultural economics generally. He and Mr. Knapp had worked in each other's counties on request before. The third was Dr. Sorghum Keynes, an extension specialist in the Department of Agricultural Economics who is assigned half-time with Extension and is available for some consultation out in the counties when this seems like the best way to use his time.

During the past few years, each of these three potential resource persons and Knapp himself, had engaged in the type of farm visits, consultation, and coaching that seemed called for in the present instance. But, Mr. Knapp wasn't very sure how well the earlier instances had worked out. He decided that he must try and find out so that the arrangement he worked out for "Corny" would be most likely to be successful. But, how should he proceed. Mr. Knapp has come to you for assistance in determining the effectiveness of past cooperative arrangements. What information would you gather about past cooperative efforts that would have an important bearing on the present problem?

# AERC EVALUATION WORKSHOP
## Case Study No. 6
## Adult Education Center Program Evaluation

Ronda Swanson is the Director of the Center for Adult Education at Downstate Community College located in a midwestern, rural-urban community. The population of this community is 82,351.

The Center for Adult Education at Downstate is one of 14 centers in the state, but the only one housed within a community college. The program at the Center primarily emphasizes adult basic education with some attention given to programs in vocational rehabilitation, occupational training, and GED certification.

Most of the funds used to support the Center's program come through the Office of the Superintendent of Public Instruction (OSPI), a state agency. OSPI receives funds from, and administers programs at the local level for, the Department of Public Aid, Title III of the 1966 Adult Education Act, and the Division of Vocational Rehabilitation. These three sources support the adult basic education and vocational rehabilitation aspects of the Center. Support for occupational training and GED programs are dependent, in some part, on tuition assessment of participants in these programs.

The total Center program has both day and evening components with a combined enrollment of approximately 305 students, both full- and part-time. Most of the participants (73%) are enrolled in the day program. Half of the students are public aid recipients, many of whom are ADC mothers. The age range of the students is 16-64 years. There are more women than men participants. Most of the participants, black and white, have low socioeconomic backgrounds. A few participants of foreign birth are enrolled in the program for the purpose of learning to read and write English.

The director of the Center is on a three-quarter time appointment and supervises both the day evening programs. The day teaching staff consists of 21 full-time and 6 part-time instructors. The evening staff includes one of the day staff members and six additional instructors. The majority of these instructors have had broad experience teaching at different levels in public schools. Only one of the staff has not had prior experience in the public schools, having been hired directly out of college.

Most of the staff members were known to the Director prior to the establishment of the Center. All candidates for instructional positions at the Center are interviewed personally by the Director at which time the philosophy and characteristics of the program are explained in depth, and observations of on-going activities are scheduled whenever possible.

In addition to the instructors, the staff includes a full-time public aid case worker and two counselors. One counselor is a rehabilitation specialist, while the other is a vocational coordinator who provides help in job placement and works with students in establishing realistic work goals.

The instructional program is diverse, being designed in many cases to meet individual student needs. For GED aspirants, there is a full-time program which runs from 9:20 to 2:20 five days a week. The course offerings in this program closely approximates the core offerings of the public high school. However, specific study requirements for students in the program take into consideration the education and experience backgrounds of the adult. To illustrate the diversity and the individualization of the program, one math teacher uses 31 different textbooks in meeting the instructional needs of the students. There are no standard beginning and ending dates for the program.

Probably the most interesting aspect of the Center's program is the adult basic education program with its unique Day Care Center program. To enhance the liklihood for potential participants (especially ADC mothers) with pre-school children to enroll in the adult basic education program a day care program has been established. This program is a cooperative venture among the Center for Adult Education, Downstate Junior College (training of para-professionals in early childhood services), and a state university (training of head-start teachers).

Due to the complexities involved in the administration and financing of the Center, the director feels accountable, in different ways, to several audiences. For example, the OSPI is primarily interested in the sequence and scope of the Center curriculum. The Department of Public Aid is most interested in student outcomes ascribed to the program. The teaching staff of the Center is interested in program revision and improvement. The director of the Center is most concerned with teacher and student recruitment and retention. The Board of Trustees of Downstate Community College views evidence of fiscal accountability as important data on which they evaluate the Center.

Given the possibility that evaluative data about the Center could be reported to at least five different audiences, how would the audience(s) for the report influence the variables that you would include in your study? To facilitate your thinking about this question a beginning set of statements have been prepared for you to consider.

Statements pertaining to possible sources of evaluative data

A T O 1. Level of achievement of participants.

A T O 2. Methods of selecting staff members.

A T O 3. Attitudes of state legislators to welfare recipients.

A T O 4. Number of contact hours between Center participants and counselors.

A T O 5. Effects on the children attending the Day Care Center while mothers participate in Center program.

A T O 6. Elimination of residence requirements for welfare eligibility.

A T O 7. Home environment of participants.

A T O 8. Number of participants obtaining employment.

A T O 9. Center budget statement for past fiscal year.

A T O 10. Description of selected instructional materials.

A T O 11. Library card applications completed by adult basic education participants.

A T O 12. Educational preparation and professional experience of teaching staff.

A T O 13. Attitudes of participants toward those culturally - different from themselves.

Judgments and standards

1. Use of national literacy norms for the general adult population.
2. Minimum standards of employers.
3. Comparison of this program with programs of the other 14 centers in the state.
4. Newspaper editorials.
5. Unemployment rate of participants as compared to non-participants.
6. Title III funding guidelines.

# APPENDIX F: Countenance Model Quiz

# COUNTENANCE MODEL QUIZ

Name ______________________________ Date ____________________

A. Circle the letter in front of the alternative that best completes the statement.

1. The main purpose of the Countenance Model is to assist evaluators who are trying to
    a. decide what things to get data on.
    b. state objectives behaviorally.
    c. correlate descriptions with judgments.
    d. develop cost-benefit ratios.

2. When plans indicate that 40% of the class time should be spent in discussion and only 25% of the time is spent that way, according to the Countenance Model, you should
    a. arrange for more discussion.
    b. change the lesson plans.
    c. gather judgment data on the discussion.
    d. correlate discussion with outcomes.

3. A teacher reports that all students have written excellent essays on "Air Pollution." According to the Countenance Model, such teacher judgments
    a. should be ignored.
    b. should be replaced by standardized tests.
    c. are valuable data.
    d. are called transactions.

4. An evaluation methodologist who urges evaluators to accept responsibility for passing judgment on a curriculum is
    a. Ralph Tyler    c. Robert Mager
    b. Lee Cronbach    d. Michael Scriven

5. In a press release Dr. Howard Benjamin, Executive Secretary of the National Schoolmasters Association, said, "Teachers should not use instructional materials that are offensive to parents." In the Countenance Model this comment is treated as
    a. hearsay.    c. gospel.
    b. a standard.    d. rationale.

6. The class is studying Gresham's Law. Several students think the lesson should provide more concrete examples. The head of the department thinks that the lesson should include fewer examples so that there is more time on the economic principle. The evaluator should
    a. help them find a compromise.
    b. report these opposite intents.
    c. do a research study on numbers of applications.
    d. get a job paying good money instead of bad.

B. Circle T if the statement is true, F is the statement is false.

7. T F Educational objectives have a prominent place in the Countenance Model.

8. T F Evaluators should limit their data to objective measurements, avoiding the collection of subjective opinions.

9. T F An important task for the evaluator is to sort all information into the three categories: antecedents, transactions, and outcomes.

10. T F Although many antecedent conditions cannot reasonably be considered the cause of student learning, information on those conditions helps make it possible for the reader of the evaluation report to decide if the findings are relevant to his school situation.

## APPENDIX G: Participant Summative Evaluation Form

# PARTICIPANT SUMMATIVE EVALUATION FORM
## AERC Evaluation Workshop

Directions: As part of our effort to evaluate the effectiveness of the AERC Evaluation Workshop, we would appreciate your completing this questionnaire. It is important that every participant complete and return this form, so that the reactions of the total group will be reflected.

We are asking you to indicate your name to facilitate coordination of returns. This questionnaire is completely confidential. Particular replies will be treated in summary form and names will *not* be associated with specific replies.

**PLEASE RETURN THIS COMPLETED FORM IN THE SELF-ADDRESSED RETURN ENVELOPE**

Name ______________________________ Date ____________________
*(last)* *(first)*

## I. WORKSHOP OBJECTIVES

1. In general, the objective of the workshop was to broaden the conceptual framework of participants from which they approach problems of evaluation in adult education. To what extent was this objective accomplished by the workshop program? *(Circle one)*

| | | | | |
|---|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |
| Very Well | Quite Well | Somewhat | Hardly | Not at all |

2. If, in general, you feel that the workshop objective was successfully achieved, indicate the *one* factor that contributed most to its success. Likewise, if you feel it was *not* successfully achieved, indicate the *one* factor that contributed most to its failure.

3. Following are listed some intended specific instructional objectives of the workshop. For each indicate the extent to which it was achieved.

| | Highly | Quite | Somewhat | Hardly | Not at all |
|---|---|---|---|---|---|
| a. To examine in detail the Stake Countenance Model of evaluation . . | ( ) | ( ) | ( ) | ( ) | ( ) |
| b. To practice using the Stake Model for the identification and categorization of variables . . . . . . . | ( ) | ( ) | ( ) | ( ) | ( ) |
| c. To design evaluation plans for typical adult education programs. . . | ( ) | ( ) | ( ) | ( ) | ( ) |
| d. To distinguish between summative and formative evaluation procedures . | ( ) | ( ) | ( ) | ( ) | ( ) |
| e. To compare and contrast research and evaluative styles of inquiry . . . | ( ) | ( ) | ( ) | ( ) | ( ) |
| f. To ascertain the role and importance of communications in evaluation . . | ( ) | ( ) | ( ) | ( ) | ( ) |

4. It is recognized that each workshop participant might have his own personal objectives for participating in the workshop. Briefly indicate what these major personal objectives were, if any.

a. ______________________________

______________________________

b. ______________________________

______________________________

c. ______________________________

______________________________

5. In general, to what extent were your personal objectives for the workshop achieved? *(Circle one)*

| | | | | |
|---|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |
| Extremely | Quite | Somewhat | Hardly | Not at all |

6. Of what importance is it for AERC to conduct, as part of its annual conference, a training workshop on a topic of relevance to adult educators?

| | | | | |
|---|---|---|---|---|
| 5 | 4 | 3 | 2 | 1 |
| Extremely Important | Quite Important | Somewhat Important | Hardly Important | Not at all Important |

## II. WORKSHOP INSTRUCTIONAL MATERIALS

1. Would you please give your opinion of the following instructional materials. *(Check one for each material.)*

| Material | Extremely Satisfactory | | Somewhat | | Not at all Satisfactory |
|---|---|---|---|---|---|
| a. Evaluation Notebook . . . | ☐ | ☐ | ☐ | ☐ | ☐ |
| b. CIRCE Attitude Scale and Profile Sheet . . . . | ☐ | ☐ | ☐ | ☐ | ☐ |
| c. Workshop Library. . . . . | ☐ | ☐ | ☐ | ☐ | ☐ |
| d. "The Interview" Film. . . | ☐ | ☐ | ☐ | ☐ | ☐ |

2. If you feel any of the above instructional materials were *less than somewhat satisfactory*, please note below any suggestions you may have for improving them.

## III. WORKSHOP SESSIONS

1. For the following indicated sessions, check (✓) the *six* sessions that impressed you most highly. Next, check the *three* with which you were least impressed.

| | Session | Highly Impressed | Least Impressed |
|---|---|---|---|
| a. | Rationale and History of Evaluation; Knox; Saturday afternoon | ( ) | ( ) |
| b. | CIRCE Attitude Scale Profile Discussion; Denny; Sunday morning | ( ) | ( ) |
| c. | Distinguishing between Research and Evaluation; Sjogren; Sunday morning | ( ) | ( ) |
| d. | Research vs. Evaluation Discussion; Staff; Sunday morning | ( ) | ( ) |
| e. | Summative and Formative Evaluation Discussion; Staff; Sunday morning | ( ) | ( ) |
| f. | Role of Evaluation Models; Gooler; Sunday morning | ( ) | ( ) |
| g. | Stake's Countenance Model of Evaluation; Stake; Sunday afternoon | ( ) | ( ) |
| h. | Discussion of Stake Model and Application to Case Study No. 6; Staff; Sunday afternoon | ( ) | ( ) |
| i. | Beyond the Countenance; Stake; Sunday evening | ( ) | ( ) |
| j. | Countenance Revisited; Denny; Sunday evening | ( ) | ( ) |
| k. | Continued Application of Stake Model to Case Study No. 6; Sjogren; Sunday evening | ( ) | ( ) |
| l. | Individual Consultation; Staff; Sunday evening | ( ) | ( ) |
| m. | Scaling; Sjogren; Monday morning | ( ) | ( ) |
| n. | Objectives; McQuarrie; Monday morning | ( ) | ( ) |

*(Continued on next page)*

| Session | Highly Impressed | Least Impressed |
|---|---|---|
| o. Item Sampling; Bunda; Monday morning . . . . | ( ) | ( ) |
| p. Unobtrusive Measures; Denny; Monday morning. . . . . . . . . . . . . . . . . . . . | ( ) | ( ) |
| q. Communications; Stake; Monday morning. . . . | ( ) | ( ) |
| r. Application of Stake Model to Case Studies Nos. 1-3; Staff; Monday morning. . . | ( ) | ( ) |
| s. Evaluation Reporting; Stake; Monday afternoon. . . . . . . . . . . . . . . . . . . | ( ) | ( ) |
| t. Panel on Issues in Evaluation; Gooler; Monday afternoon . . . . . . . . . . . . . . . | ( ) | ( ) |

2. For those sessions with which you were *highly impressed,* please specify briefly any aspects of the subject or discussion topic, the methods, or the staff that impressed you highly.

3. For those sessions with which you were *least impressed,* please specify briefly any aspects of the subject or discussion topic, the methods, or the staff that may have most improved the sessions.

## IV. WORKSHOP INSTRUCTIONAL STAFF

1. Indicate the extent to which the instructional staff was *enthusiastic* about the topics they presented.

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Very | Quite | Somewhat | Hardly | Not at all |

2. Indicate the extent to which the staff was properly *prepared.*

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Very | Quite | Somewhat | Hardly | Not at all |

3. Indicate the degree to which the staff was *helpful and friendly.*

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Very | Quite | Somewhat | Hardly | Not at all |

4. In general, rate the instructional staff.

Superior . . . . . ( )
Good . . . . . . . ( )
Average. . . . . . ( )
Fair . . . . . . . ( )
Poor . . . . . . . ( )

## V. WORKSHOP FACILITIES

1. There are many parts of a workshop experience that can either contribute to your satisfaction or detract from it. For each of the following, would you indicate how satisfied you have been.

a. Hotel Rooms

Really outstanding . . . . . ( )
Quite satisfactory . . . . . ( )
Average. . . . . . . . . . . ( )
Just acceptable. . . . . . . ( )
Inadequate . . . . . . . . . ( )

b. Meeting Rooms

Really outstanding . . . . . ( )
Quite satisfactory . . . . . ( )
Average. . . . . . . . . . . ( )
Just acceptable. . . . . . . ( )
Inadequate . . . . . . . . . ( )

c. Eating Facilities

Really outstanding . . . . . ( )
Quite satisfactory . . . . . ( )
Average. . . . . . . . . . . ( )
Just acceptable. . . . . . . ( )
Inadequate . . . . . . . . . ( )

2. In general how do you rate the facilities?

Really outstanding . . . . . ( )
Quite satisfactory . . . . . ( )
Average. . . . . . . . . . . ( )
Just acceptable. . . . . . . ( )
Inadequate . . . . . . . . . ( )

## VI. WORKSHOP OUTCOMES

1. Now that you have had time to reflect on what was presented at the workshop, have you on the most part

reverted back to the ideas you had about evaluation prior to attending the workshop . . . ( )

maintained the ideas presented in the workshop . . . . . . . . . . . . . . . . . . . . . ( )

built upon what you gained in the workshop . . . ( )

2. Since participating in the workshop, to what extent do you feel more competent to approach and conduct evaluation studies?

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Highly | Quite | Somewhat | Hardly | Not at all |

3. To what extent have you used what you have gained in the workshop?

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Substantially | | Moderately | | Not at all |

4. Have you read or discussed with anyone various aspects of or about evaluation since participating in the workshop?

Yes . . . ( ) No . . . ( )

5. *(If yes)* Briefly describe the nature of your activity.

__________________________________________________________
__________________________________________________________
__________________________________________________________

6. How would you rate the AERC Evaluation Workshop with similar workshops you have attended?

Substantially better . . . . ( )
About the same . . . . . . . ( )
Substantially worse. . . . . ( )

7. How would you rate the AERC Evaluation Workshop with workshops you have conducted?

Substantially better . . . . ( )
About the same . . . . . . . ( )
Substantially worse. . . . . ( )
Never ran a workshop . . . . ( )

8. In general, how much impact do you think the workshop had on the field?

| 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|
| Great Impact | Much Impact | Some Impact | Little Impact | No Impact |

Thank you for your cooperation. Please return this completed form in the self-addressed return envelope to: Arden Grotelueschen, CIRCE, 270 Education Building, University of Illinois, Urbana, Illinois 61801.

# APPENDIX H: AERC Evaluation Workshop Participant Information

## AERC EVALUATION WORKSHOP
### Participant Information

1. Name (Dr., Mr., Mrs., Miss) ______________________________________

   *(last)* *(first)*

2. Indicate highest degree, major, institution and year degree received.

   Bachelor's ( ) Master's ( ) Doctorate ( )

   Major ____________ Institution ______________________ Year ______

3. Indicate what percent of your time is allotted to each of the following:

   Research and Evaluation . . . _____%
   Teaching. . . . . . . . . . . _____%
   Administration. . . . . . . . _____%
   Service . . . . . . . . . . . _____%
   100 %

4. Describe briefly the nature and specific duties of those aspects of your present job that relate to evaluation. If none, please indicate so.

   ______________________________________________________________

   ______________________________________________________________

   ______________________________________________________________

5. Did you attend the AERC Paper Session? Yes ( ) No ( )

6. How much have you looked at the black Evaluation Workshop notebook?

   Haven't looked at it . . . . . . . ( )
   Thumbed through it . . . . . . . . ( )
   Skimmed some of the contents . . . ( )
   Read some of the contents. . . . . ( )

   7. *(If you read some of the contents)* What did you read? Give article title or author name.

      ___________________________________________________________

      ___________________________________________________________

      ___________________________________________________________

8. Did the Evaluation Notebook

   Turn you on . . . . . . ( )
   Do nothing for you. . . ( )
   Turn you off. . . . . . ( )

9. For each of the following evaluation topics indicate the extent of your familiarity.

| | Highly Familiar | Somewhat Familiar | Not Familiar |
|---|---|---|---|
| Rationale and history of evaluation . . . . . . . . . . . . . . . | ( ) | ( ) | ( ) |
| Evaluation models. . . . . . . . . . . | ( ) | ( ) | ( ) |
| Distinction between research and evaluation . . . . . . . . . . . . . . . | ( ) | ( ) | ( ) |
| Summative and formative evaluation . | ( ) | ( ) | ( ) |
| Stake's Countenance Model. . . . . . | ( ) | ( ) | ( ) |
| Communications . . . . . . . . . . . . | ( ) | ( ) | ( ) |
| Scaling techniques . . . . . . . . . . | ( ) | ( ) | ( ) |
| Educational objectives . . . . . . . . | ( ) | ( ) | ( ) |
| Item sampling. . . . . . . . . . . . . | ( ) | ( ) | ( ) |
| Unobtrusive measures . . . . . . . . . | ( ) | ( ) | ( ) |
| Evaluation reporting . . . . . . . . . | ( ) | ( ) | ( ) |

# APPENDIX I: 1970 Adult Education Research Conference Participant Roster

# 1970 ADULT EDUCATION RESEARCH CONFERENCE

## Participant Roster

Mr. Albert Adams
State Superintendent of
Adult Education
State Department of Education
600 Wyndhurst
Baltimore, Maryland 21210

Mr. Pierre Amyot
Director of Research
Service D;Éducation Permanente
Université De Montréal
Montréal, Canada

Mr. James Anderson
Room 479
1414 East 59th Street
University of Chicago
Chicago, Illinois

Dr. Irene Beavers
Associate Professor
Iowa State University
166C Mackay
Ames, Iowa

Mr. Paul Belanger
Research Officer
Institut Canadian Education
Adultes
506 est Ste-Catherine
Montreal, Canada

Mr. Ian Bell
Lecturer (Communications)
Centralia College of Agricultural
Technology
P. O. Huron Park
Ontario, Canada

Mr. Jack Blaney
Associate Director
Extension Department
University of British Columbia
Vancouver, Canada

Mr. Patrick Borich
District Supervisor
County Extension Work
University of Minnesota
St. Paul, Minnesota 56101

Dr. Milton Boyce
Research Associate
North Carolina State University
109 Ricks Hall
Raleigh, North Carolina 27607

Dr. Emory Brown
Assistant Director, Cooperative
Extension
Penn State University
University Park, Pennsylvania 16802

Dr. Robert L. Bruce
Associate Professor of Extension
Education
Cornell University
Stone Hall
Ithaca, New York 14850

Mr. Darcie Byrn
Extension Research Specialist
Federal Extension Service
U. S. Department of Agriculture
Washington, D. C. 20250

Mr. James Camp
Research Assistant
North Carolina State University
109 Ricks Hall
Raleigh, North Carolina 27607

Dr. Eleanor K. Caplan
Research Director
Cleveland College
Case Western Reserve University
Baker Building
Cleveland, Ohio 44106

Dr. Doris Chertow
Editor, Publications in Continuing
Education
Syracuse University
105 Roney Lane
Syracuse, New York 13210

Dr. Louise Colley
Assistant Professor
Coordinator of Continuing Education
University of Guelph
Guelph, Ontario, Canada

Mr. Lin Compton
Community Adult Education
University of Michigan
Ann Arbor, Michigan

Miss Constance Cooper
International House
Room 527
1414 East 59th Street
University of Chicago
Chicago, Illinois 60637

Dr. Harlan Copeland
Associate Professor
Syracuse University
Syracuse, New York 13210

Mrs. Reba Davis
352 Education Building
University of Illinois
Urbana, Illinois 61801

Mr. Derek Dawson
Director of Staffing
Training Division
Post Office Department H.Q.
Ottawa 8, Ontario, Canada

Dr. Ralph Dobbs
Major Advisor in Adult Education
University of Missouri
Columbia, Missouri 65201

Dr. Robert Dolan
Professor of Adult Education
North Carolina State University
113 Ricks Hall
Raleigh, North Carolina 27607

Dr. James Draper
Associate Professor
Ontario Institute for Studies
in Education
252 Bloor Street, West
Toronto, Canada

Dr. Donnie Dutton
Director of Adult Education
Memphis State University
Memphis, Tennessee

Miss Elizabeth Elliott
Program Leader, Family Living
Program
University Extension
University of Wisconsin
501 Extension Building
Madison, Wisconsin

Dr. Jack C. Everly
Coordinator of Instructional Resources
College of Agriculture
University of Illinois
69 Mumford Hall
Urbana, Illinois 61801

Mrs. Ann Fales
5331 South Harper
University of Chicago
Chicago, Illinois 60615

Dr. James Farmer
Assistant Professor of Adult Education
University of California
Los Angeles, California 90024

Dr. Glen Farrell
Professor of Continuing Education
University of Saskatoon
Saskatoon, Canada

Dr. Josephine Flaherty
Assistant Professor
Ontario Institute for Studies in
Education
252 Bloor Street, West
Toronto 5, Canada

Mr. Laverne Forest
University of Wisconsin
504 H Eagle Heights
Madison, Wisconsin 53705

Miss Doris Frazier
West 419 Owen Hall
Michigan State University
East Lansing, Michigan

Habeeb Ghatala
Associate Director
Humboldt State College
University of Wisconsin
Madison, Wisconsin

Dr. Betty Giuliani
Director, Continuing Education Research
Michigan State University
East Lansing, Michigan

Mr. Ralph Goodman
Supervisor
State of Illinois
316 South 2nd Street
Springfield, Illinois

Mr. Stanley Grabowski
Assistant Director
ERIC/AE
Syracuse, New York 13210

Miss Virginia Griffin
Lecturer
Ontario Institute for Studies in Education
102 Bloor Street, West
Toronto 5, Ontario, Cànada

Dr. John Gross
Extension Studies Specialist
University of Missouri
Columbia, Missouri 65201

Mr. Arden Grotelueschen
Assistant Professor
CIRCE, 270 Education Building
University of Illinois
Urbana, Illinois 61801

Miss Yolanda Hall
Graduate Student
University of Chicago
Chicago, Illinois

Mrs. Donna Ilyin
Teacher and Consultant
San Francisco Unified School District
76 Sixth Avenue
San Francisco, California 94118

Dr. Irene Imbler
Assistant Professor of Adult Education
College of Human Ecology
Cornell University
Ithaca, New York

Dr. Roy Ingham
Associate Professor
Florida State University
Tallahassee, Florida

Dr. Clara Kanun
Director of Research
305 Nolte Center
University of Minnesota
Minneapolis, Minnesota 55414

Dr. Eldora Keske
Associate Professor, Youth Development Specialist
University of Wisconsin Extension
606 State Street
Madison, Wisconsin 53706

Mr. Russell Kleis
Director, Graduate Studies in Continuing Education
Michigan State University
East Lansing, Michigan

Dr. Alan Knox
Professor
Teachers College
Columbia University
New York, New York 10027

Dr. Charles Kozoll
Associate Project Director
Southern Regional Education Board
130 Sixth Street, N.W.
Atlanta, Georgia 30313

Mrs. Penny Landvogt
Extension Home Economist
University of Wisconsin
Rock County Court House
Janesville, Wisconsin

Mr. Ivan Lappin
Field Services
National Congress of Parents and Teachers
700 North Rush
Chicago, Illinois

Mr. George Lehmann
Assistant to the Dean
Adult and Continuing Education
University of Toledo
Toledo, Ohio

Dr. Walter A. Leirman
Associate Professor
Leuven University
2839 Revelstoke Court
Vancouver 8, Canada

Mrs. Eileen E. Lineberry
Director of Continuing Education
for Women
University of Missouri
210 Whitten Hall
Columbia, Missouri 65201

Mr. Paul Lineberry
Program Analyst
Missouri Law Enforcement
Assistance
500 Jefferson Building
Jefferson City, Missouri

Dr. Ann Litchfield
Assistant Professor
University of Chicago
5835 South Kimbark
Chicago, Illinois 60637

Dr. Huey B. Long
Associate Professor of
Adult Education
University of Georgia
Athens, Georgia

Mr. E. J. Lueder
University of Wisconsin
Madison, Wisconsin

Mr. James Martin
University of Wisconsin
1415 University Avenue
Madison, Wisconsin

Dr. Homer Mazer
County Agriculture Agent
Pennsylvania State University
Clearfield, Pennsylvania

Mrs. Molly McCampbell
Research Assistant/Program
Coordinator
Kansas City School District
1903 Paseo
Kansas City, Missouri 64108

Mr. Howard Y. McClusky
Professor of Educational Psychology
and Consultant in Community
Adult Education
University of Michigan
Ann Arbor, Michigan

Dr. Allen Menlo
Professor of Education
University of Michigan
Ann Arbor, Michigan

Dr. Jack Mezirow
Associate Professor of Adult Education
Teacher College
Columbia University
New York, New York 10027

Dr. Mason Miller
Director, Institute for Extension
Personnel Development
Michigan State University
410 Ag Hall
East Lansing, Michigan 48823

Mr. Gary Moss
Research Assistant
Memphis State University
Memphis, Tennessee

Miss Gwenna Moss
Instructor
University of Wisconsin
Madison, Wisconsin

Mr. Jim Nichols
Administrative Intern
University of Toledo
3607 Burton Avenue
Toledo, Ohio

Mr. Paul Norton
Extension Director
Age Center for New England
73 Newberry Street
Boston, Massachusetts

Mr. Albert Parsons
Director of Extension
Ryerson Polytechnical Institute
50 Gould Street
Toronto, Canada

Dr. John Peters
Assistant Professor of Adult Education
University of Tennessee
Knoxville, Tennessee

Dr. Bernadine H. Peterson
Family Living Program Leader
University Extension
University of Wisconsin
Madison, Wisconsin 53706

Dr. Lee Porter
Program Administrator for Liberals Arts and Director of Research
Syracuse University
Syracuse, New York 13204

Dr. William H. Puder
Research Associate
Center for Occupational Education
North Carolina State University
Raleigh, North Carolina

Mr. Louis Reeves
Research Consultant
Public Service Commission
Ottawa, Ontario, Canada

Dr. Russell D. Robinson
Associate Professor
University of Wisconsin
Milwaukee, Wisconsin

Dr. Harold Rose
Research Administrator
Appalachian Adult Basic Education Demonstration Center
Morehead State University
Morehead, Kentucky 40351

Mr. Roy H. Ryan
Director, Middle Adult Ministries
United Methodist Board of Education
P. O. Box 871
Nashville, Tennessee 37202

Dr. Don F. Seaman
Assistant Professor of Adult Educ.
Mississippi State University
State College, Mississippi 39759

Dr. Ronald Shearon
Assistant Professor
North Carolina State University
Raleigh, North Carolina

Mrs. Alvaloise Simms
Head Teacher
Kansas City Schools
Kansas City, Missouri

Dr. Robert M. Smith
Associate Director
University Center of Adult Education
Detroit, Michigan

Dr. Robert Snyder
Assistant Professor and Coordinator
Adult Education Program
University of South Carolina
Columbia, South Carolina 29208

Mrs. Hazel Taylor Spitze
Associate Professor
University of Illinois
Urbana, Illinois 61801

Miss Vida Stanius
Research Associate
Ontario Institute for Studies in Education
Toronto 5, Ontario, Canada

Dr. Sara Steele
Associate Professor
University of Wisconsin
Milwaukee, Wisconsin

Mr. Ronald Szczypkowski
Research Associate
Milbank Memorial Fund
40 Wall Street
New York, New York

Dr. D. Robert Troyer
Assistant Director
Cooperative Extension
Purdue University
Lafayette, Indiana

Dr. Clive Veri
Assistant Dean
University of Missouri
8001 Natural Bridge
St. Louis, Missouri 63121

Dr. Ted Warren
Minister
Lawton Street Christian Church
416 South Lawton
Jasonville, Indiana

Mr. Harold Wecke
Head: Special Programs and Research
Division of University Extension
University of Illinois
120 Illini Hall
Champaign, Illinois 61820

Mr. Harry Zimmerman
Associate Professor
University of Wisconsin
University Extension
606 State Street
Madison, Wisconsin

# APPENDIX J: AERC Evaluation Workshop Schedule

## AERC EVALUATION WORKSHOP SCHEDULE

| Day/Time | Room | Session | Leader | Topic/Activity |
|---|---|---|---|---|
| **Saturday Afternoon** | | | | |
| 3:20 - 3:45 | Starlite | Maxi (total group) | Grotelueschen, Staff | Introduction, administration of pre-instrument, material distribution, and plans and procedures |
| 3:45 - 4:30 | Starlite | Maxi | Knox | Rationale and history of evaluation |
| 4:30 - 4:50 | Starlite | Maxi | Grotelueschen | Overview of Stake countenance model |
| 4:50 - 5:00 | Starlite | Maxi | Grotelueschen, Staff | Pre-administration of CIRCE Attitude Scale |
| **Saturday Evening** | [Participants responsible for scoring CIRCE Attitude Scale and completing individual profile sheets. Highly desirable for participants to scan pages 9 - 42 (Glass article), and essential for participants to read pages 43 - 60 (Stake article) in Evaluation Workshop Notebook.] | | | |
| **Sunday Morning** | | | | |
| 8:30 - 9:15 | Starlite | Maxi | Denny | Discussion of CIRCE Attitude Scale profiles |
| 9:15 - 9:35 | Starlite | Maxi | Sjogren | Distinguishing between research and evaluation |
| 9:35 - 10:00 | Starlite | Midi (1/3 of group-assigned) | Denny | Discussion of research vs. evaluation |
| | Rm. 8-9 | Midi | McQuarrie | |
| | Rm. 6-7 | Midi | Grotelueschen | |
| 10:00 - 10:20 | (Coffee in Starlite Foyer) | | | |
| 10:20 - 11:00 | Starlite | Midi | Denny | Summative and formative evaluation |
| | Rm. 8-9 | Midi | Sjogren | |
| | Rm. 6-7 | Midi | Grotelueschen | |

SCHEDULE CONTINUED . . .

| Day/Time | Room | Session | Leader | Topic/Activity |
|---|---|---|---|---|
| 11:00 - 11:45 | Starlite | Maxi | Gooler | Role of evaluation models |
| 11:45 - 12:00 | Starlite | Maxi | Staff | Workshop formative evaluation No. 1 |
| Sunday Afternoon | | | | |
| 12:00 - 1:30 | (Lunch break) | | | |
| 1:30 - 3:00 | Starlite | Maxi | Stake | Stake's Countenance Model of Evaluation |
| 3:00 - 3:20 | (Coffee in Starlite Foyer) | | | |
| 3:20 - 4:00 | Starlite | Midi | Stake | Discussion of Stake model |
| | Rm. 8-9 | Midi | Denny | |
| | Rm. 6-7 | Midi | Sjogren | |
| 4:00 - 4:45 | Starlite | Midi | Stake | Application of Stake model to Case Study No. 6 |
| | Rm. 8-9 | Midi | Denny | |
| | Rm. 6-7 | Midi | Sjogren | |
| 4:45 - 5:00 | Starlite | Midi | Staff | Workshop formative evaluation No. 2 |
| | Rm. 8-9 | Midi | Staff | |
| | Rm. 6-7 | Midi | Staff | |
| 5:00 - 7:30 | (Dinner break) | | | |

SCHEDULE CONTINUED . . .

| Day/Time | Room | Session | Leader | Topic/Activity |
|---|---|---|---|---|
| **Sunday Evening** | | | | |
| 7:30 - 8:30 | Starlite | Midi-Mini (voluntary) | Stake | Beyond the countenance |
| | Rm. 8-9 | Midi-Mini | Denny | Countenance revisited |
| | Rm. 6-7 | Midi-Mini | Sjogren | Continued application of Stake model to Case Study No. 6 |
| 8:30 - 9:30 | (Individual consultations by workshop staff with participants on special evaluation problems.) | | | |
| **Monday Morning** | | | | |
| 8:40 - 10:00 (Two 40 minute repeat performances -- participant's choice.) | Rm. 9 | Mini | Sjogren | Scaling |
| | Rm. 8 | Mini | McQuarrie | Objectives |
| | Rm. 7 | Mini | Bunda | Item Sampling |
| | Rm. 6 | Mini | Denny | Unobtrusives |
| 10:00 - 10:20 | (Coffee in Starlite Foyer) | | | |
| 10:20 - 11:00 | Starlite | Maxi | Stake | Communications |
| 11:00 - 12:00 | Starlite | Midi | Grotelueschen | Application of Stake model to Case No. 1 |
| | Rm. 8-9 | Midi | Denny | Application of Stake model to Case No. 2 |
| | Rm. 6-7 | Midi | Knox | Application of Stake model to Case No. 3 |

## SCHEDULE CONTINUED . . .

| Day/Time | Room | Session | Leader | Topic/Activity |
|---|---|---|---|---|
| Monday Afternoon | | | | |
| 12:00 - 1:30 | (Lunch break) | | | |
| 1:30 - 2:15 | Starlite | Maxi | Stake | Evaluation reporting |
| 2:15 - 2:45 | Starlite | Maxi | Gooler (panel moderator) | Issues in evaluation |
| 2:45 - 3:00 | Starlite | Maxi | Grotelueschen | Workshop wrap-up, administration of post-instruments |

# APPENDIX K: AERC Evaluation Workshop Instructional Staff

## AERC EVALUATION WORKSHOP
### Instructional Staff

Mary Anne Bunda
Research Assistant,
*CIRCE;
Ph.D. candidate with interests in statistics, design, measurement and evaluation.

Terry Denny
Director, Educational Products Information Exchange, Office of Evaluation Studies, CIRCE;
interests in evaluation field study, site visitations, interviewing, unobtrusive measures, and table tennis.

Dennis Gooler
Research Assistant,
CIRCE;
Ph.D. candidate with interests in curriculum evaluation and evaluating evaluation.

Arden Grotelueschen
Assistant Professor, CIRCE;
interests in adult learning and instruction, measurement and evaluation, and land speculation.

Alan B. Knox
Professor, Teachers College, Columbia University;
interests in about everything having to do with adult education, and painting, and music composition.

Duncan McQuarrie
Research Assistant,
CIRCE;
Ed.D. candidate with interests in evaluation, measurement, learning, and new cars (at the moment).

Marge Pjojian
Research Assistant,
CIRCE;
Ed.D. candidate with interests in evaluation and elgible men.

Doug Sjogren
Professor, CIRCE;
interests in measurement and evaluation, and trout fishing.

Bob Stake
Professor and Associate Director,
CIRCE;
interests in curriculum evaluation, psychometrics, and geneology.

Gary Storm
Research Assistant,
CIRCE;
doctoral student interested in evaluation, ecological studies, leisure, and summer camps.

*Center for Instructional Research and Curriculum Evaluation, College of Education, University of Illinois, Urbana, Illinois 61801.

# APPENDIX L: 1970 AERC Evaluation Workshop Observation Guide

## 1970 ADULT EDUCATION RESEARCH CONFERENCE (AERC)
### Evaluation Workshop Observation Guide

1.00 Session and Observer

1.10 Session Identification ________________________________________

1.20 Observer ________________________________________________

2.00 Session Characteristics

2.10 Attendance at beginning of session ____________________________

2.20 Starting time _____________________________________________

2.30 Ending time ______________________________________________

2.40 Attendance at end of session __________%

3.00 Adequacy of meeting room

3.10 Seating: Adequate ( ) Inadequate ( )
3.20 Accoustics: Very Good ( ) Acceptable ( ) Poor ( )
3.30 Lighting: Very Good ( ) Acceptable ( ) Poor ( )
3.40 Audio-visual: Very Good ( ) Acceptable ( ) Poor ( )
3.50 General instructional climate (Note only undesirable conditions):

4.00 Observer notes of session activity (Topics covered, audience reactions, general overview):

5.00 Rating of audience

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 5.10 | 1<br>Inattentive | 2 | 3 | 4 | 5 | 6 | 7<br>Attentive |
| 5.20 | 1<br>Hostile | 2 | 3 | 4 | 5 | 6 | 7<br>Friendly |
| 5.30 | 1<br>Restless | 2 | 3 | 4 | 5 | 6 | 7<br>Restful |
| 5.40 | 1<br>Disinterested | 2 | 3 | 4 | 5 | 6 | 7<br>Interested |

6.00 Notes derived from interacting with participants. (Look for understandings; note general attitude; what did they talk about; level of understandings.)

6.10 Thumbnail sketch of person or persons talked to:

6.20 Notes on interactions.

ERIC Clearinghouse
APR 1 9 1971
on Adult Education